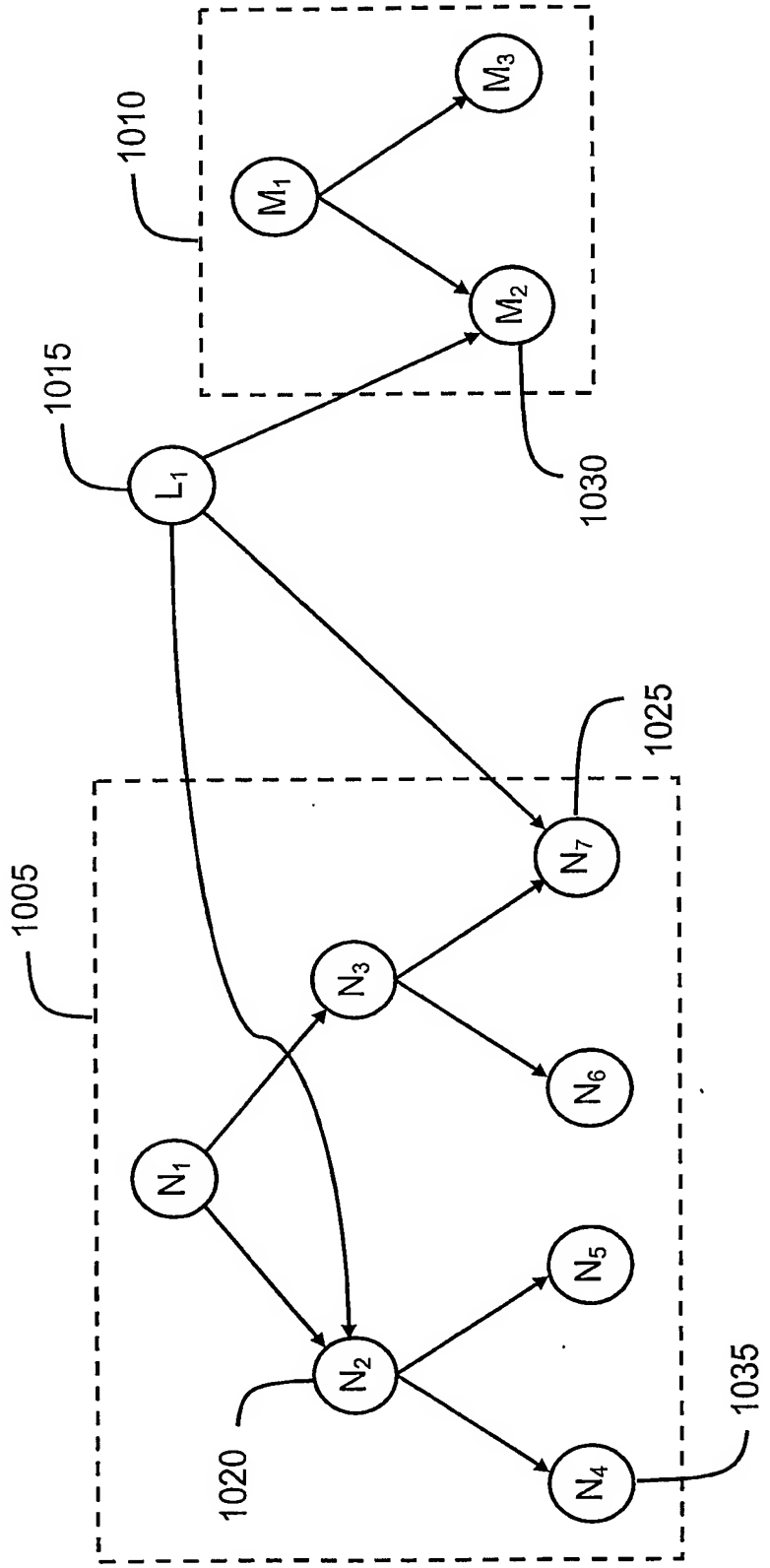
1010
M1
M3
M2
1030
1015
L1
1025
N7
1005
N3
N6
N1
N5
N2
1020
N4
1035

Fig. 1

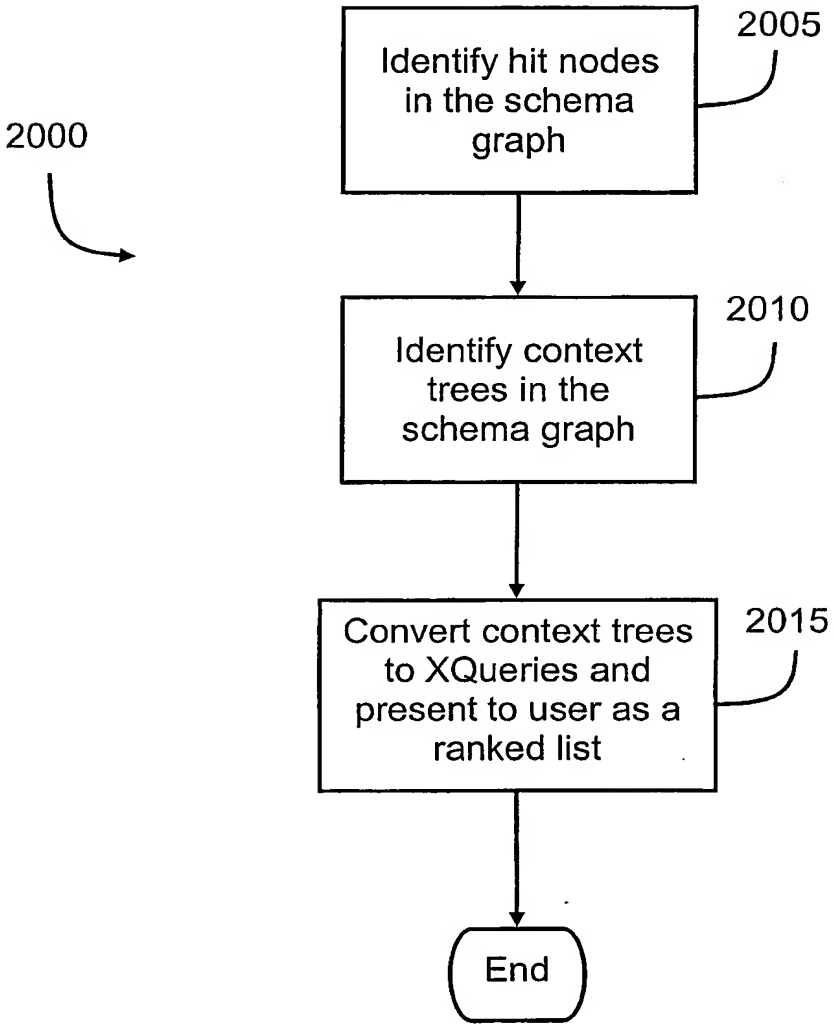
2000
Identify hit nodes in the schema graph
2005
Identify context trees in the schema graph
2010
Convert context trees to XQueries and present to user as a ranked list
2015
End

Fig. 2

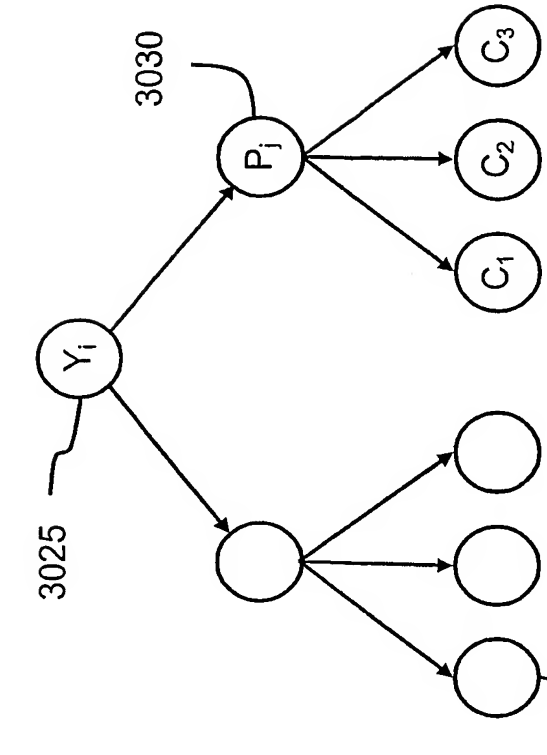
3025
Y_i
3030
P_j
C_3
C_2
C_1
3035
X
General case

Fig. 3B

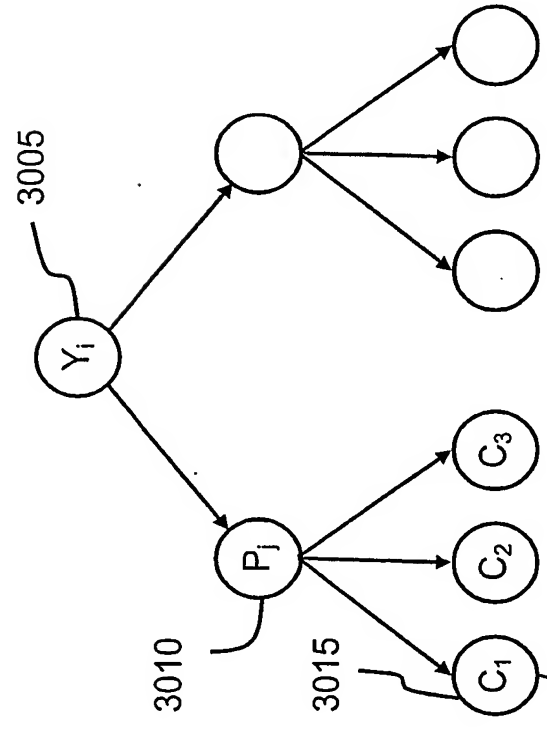
3005
Y_i
3010
P_j
3015
C_1
C_2
C_3
3020
X
P_j lying along path to hit node

Fig. 3A

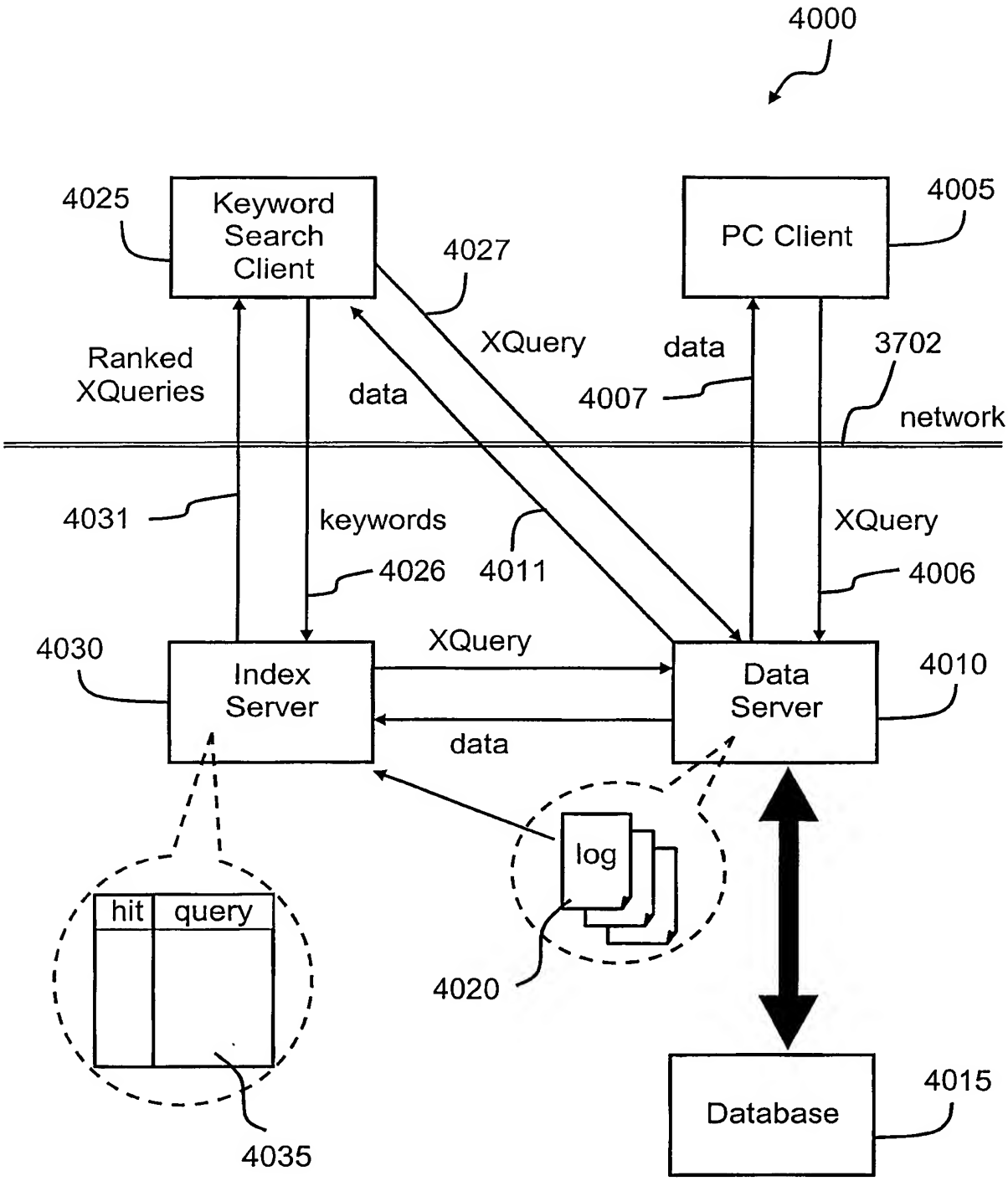
4000
4025
Keyword Search Client
4027
PC Client
4005
XQuery
data
3702
Ranked XQueries
data
4007
network
4031
keywords
4026
4011
XQuery
4006
4030
Index Server
XQuery
Data Server
4010
data
log
hit
query
4020
4035
Database
4015

Fig. 4

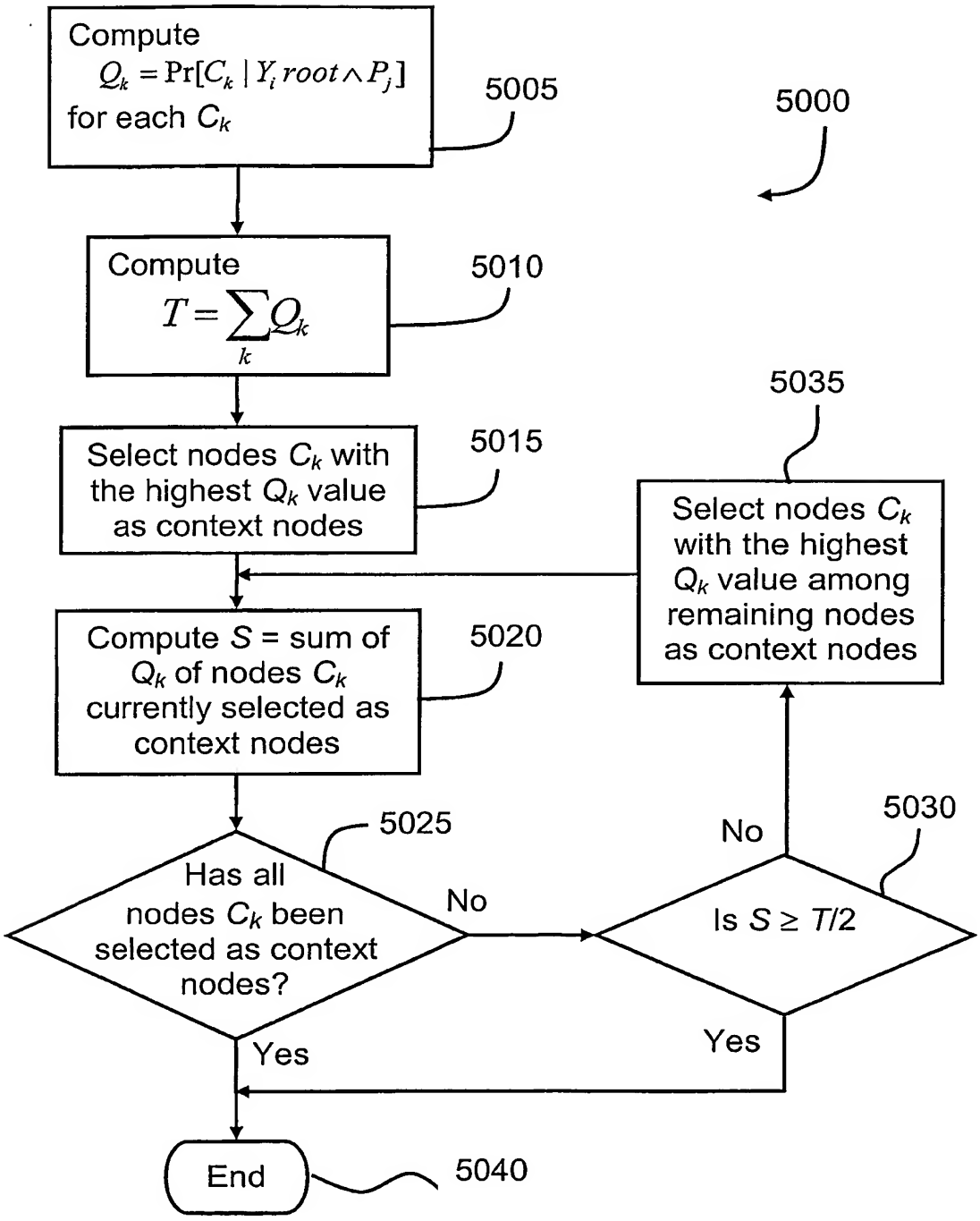
Compute $Q_k = \Pr[C_k \mid Y_i\, root \wedge P_j]$ for each C_k
5005
5000
Compute $T = \sum_k Q_k$
5010
Select nodes C_k with the highest Q_k value as context nodes
5015
5035
Select nodes C_k with the highest Q_k value among remaining nodes as context nodes
Compute S = sum of Q_k of nodes C_k currently selected as context nodes
5020
5025
Has all nodes C_k been selected as context nodes?
No
5030
No
Is $S \geq T/2$
Yes
Yes
End
5040

Fig. 5

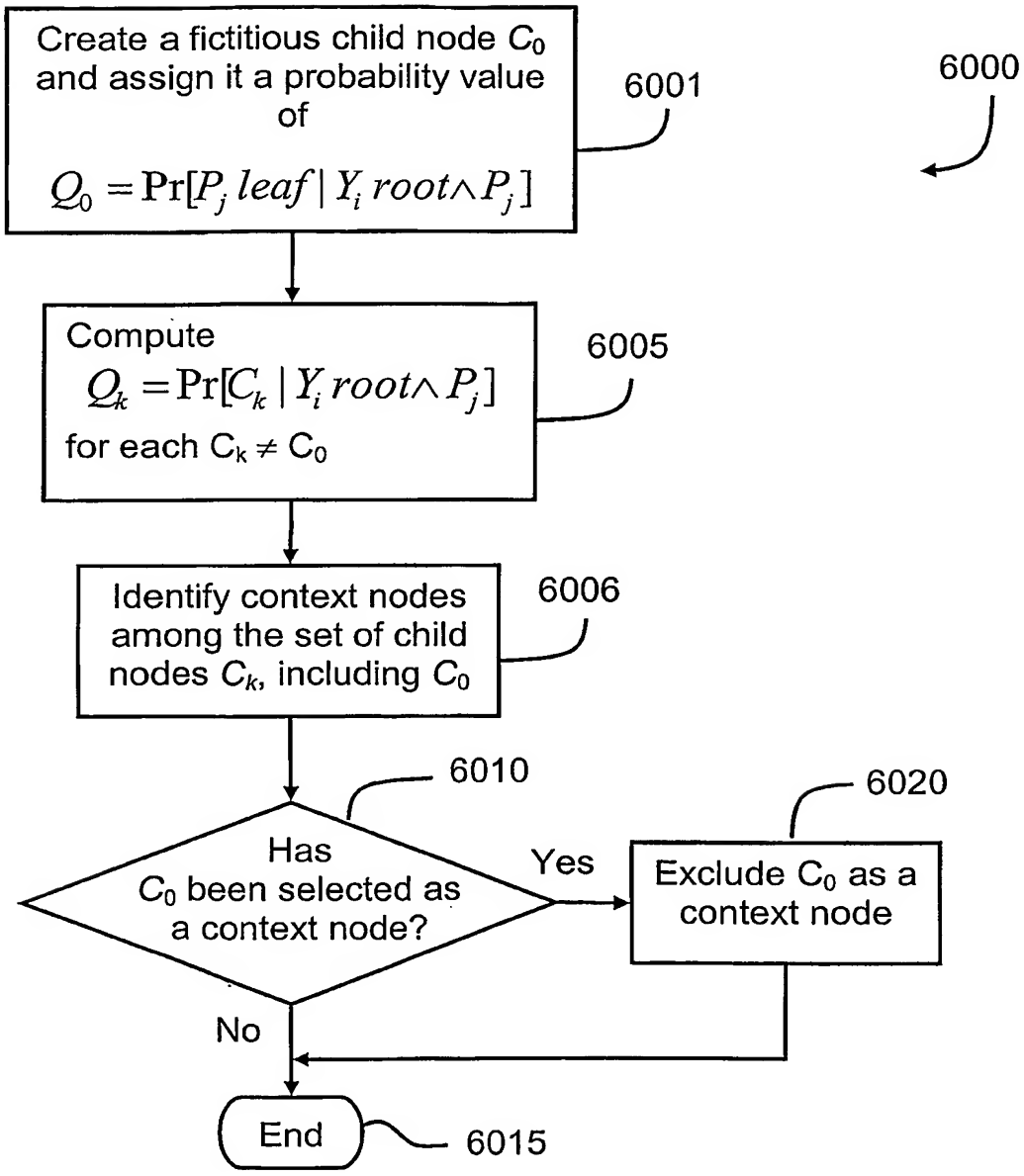
Create a fictitious child node C_0 and assign it a probability value of
$Q_0 = \Pr[P_j \, leaf \mid Y_i \, root \wedge P_j]$
6001
6000
Compute
$Q_k = \Pr[C_k \mid Y_i \, root \wedge P_j]$
for each $C_k \neq C_0$
6005
Identify context nodes among the set of child nodes C_k, including C_0
6006
6010
Has C_0 been selected as a context node?
Yes
6020
Exclude C_0 as a context node
No
End
6015

Fig. 6

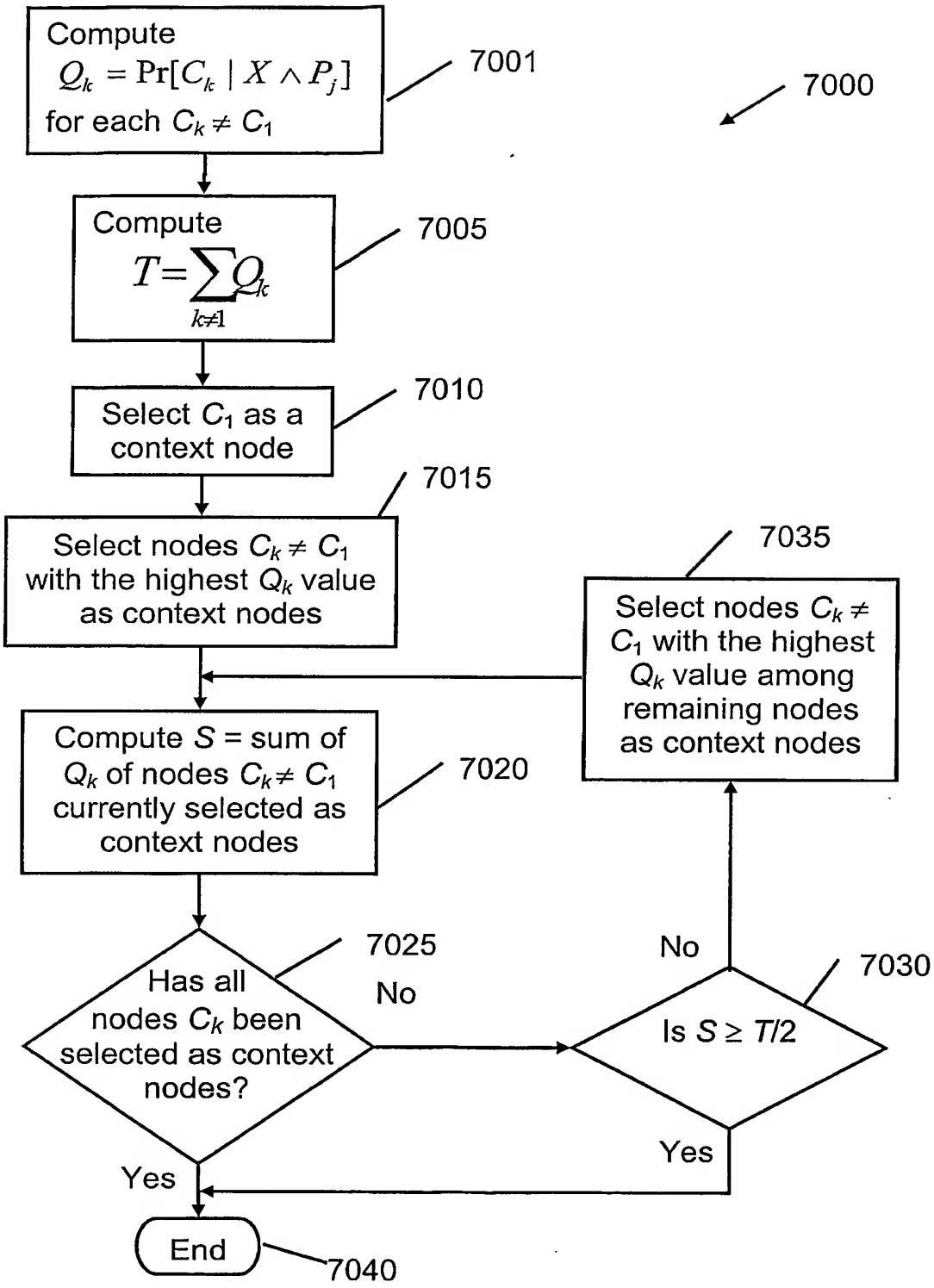
Compute $Q_k = \Pr[C_k \mid X \wedge P_j]$ for each $C_k \neq C_1$
7001
7000
Compute $T = \sum_{k \neq 1} Q_k$
7005
Select C_1 as a context node
7010
7015
Select nodes $C_k \neq C_1$ with the highest Q_k value as context nodes
7035
Select nodes $C_k \neq C_1$ with the highest Q_k value among remaining nodes as context nodes
Compute S = sum of Q_k of nodes $C_k \neq C_1$ currently selected as context nodes
7020
7025
Has all nodes C_k been selected as context nodes?
No
No
7030
Is $S \geq T/2$
Yes
Yes
End
7040

Fig. 7

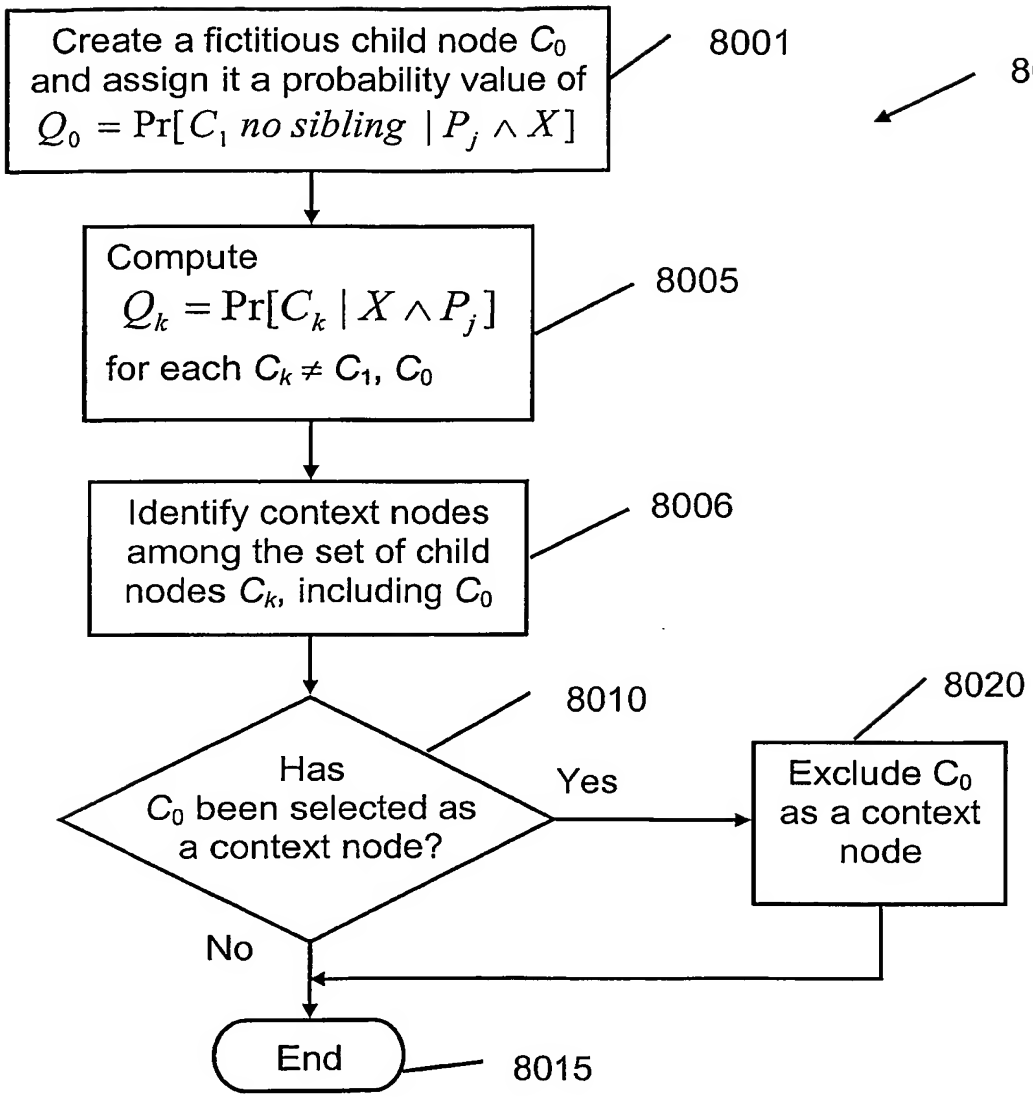
Create a fictitious child node C_0 and assign it a probability value of $Q_0 = \Pr[C_1 \text{ no sibling } | P_j \wedge X]$
8001
8000
Compute $Q_k = \Pr[C_k | X \wedge P_j]$ for each $C_k \neq C_1, C_0$
8005
Identify context nodes among the set of child nodes C_k, including C_0
8006
Has C_0 been selected as a context node?
8010
Yes
Exclude C_0 as a context node
8020
No
End
8015

Fig. 8

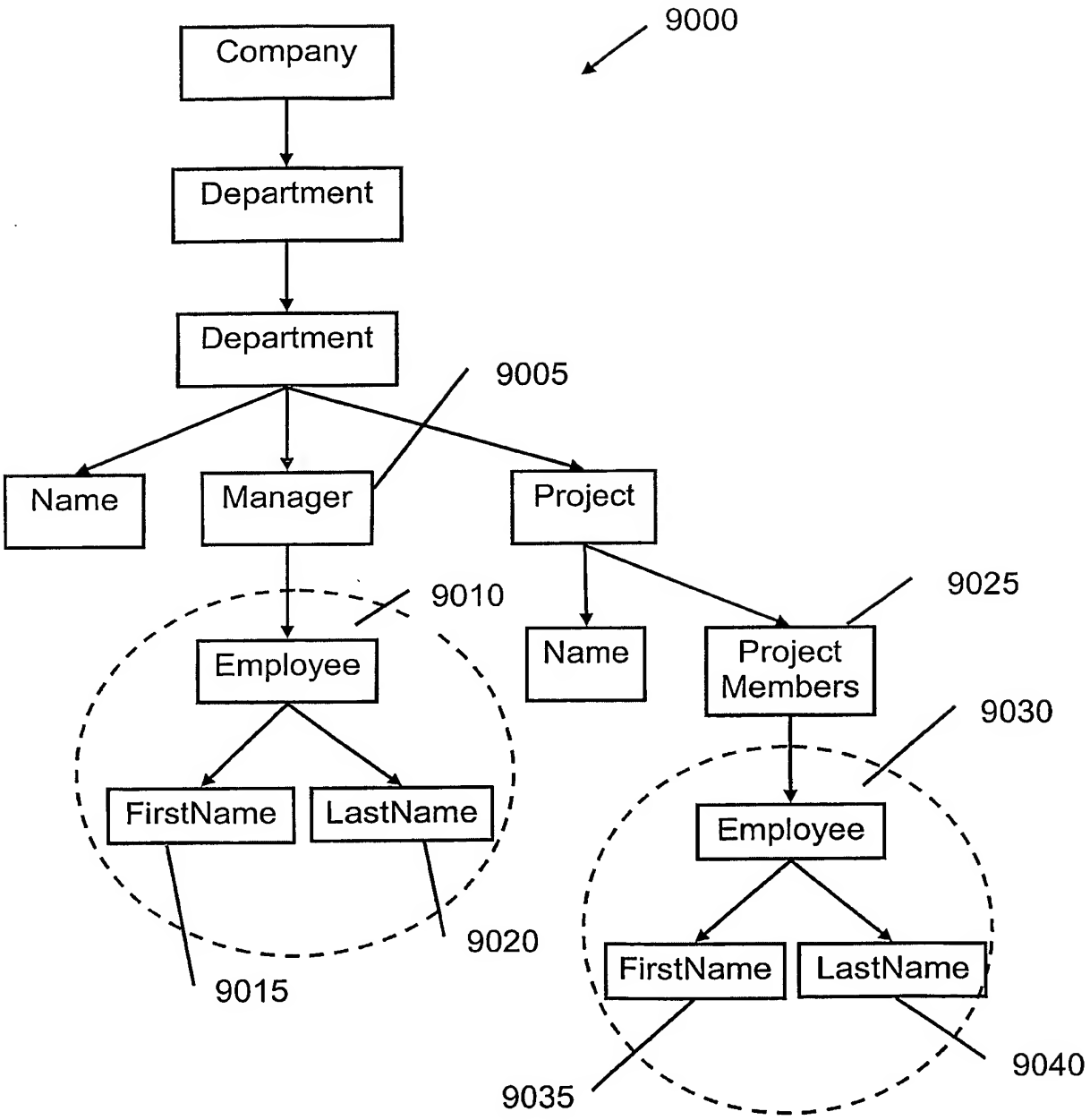
9000
Company
Department
Department
9005
Name
Manager
Project
9010
9025
Employee
Name
Project Members
9030
FirstName
LastName
Employee
9020
9015
FirstName
LastName
9035
9040

Fig. 9

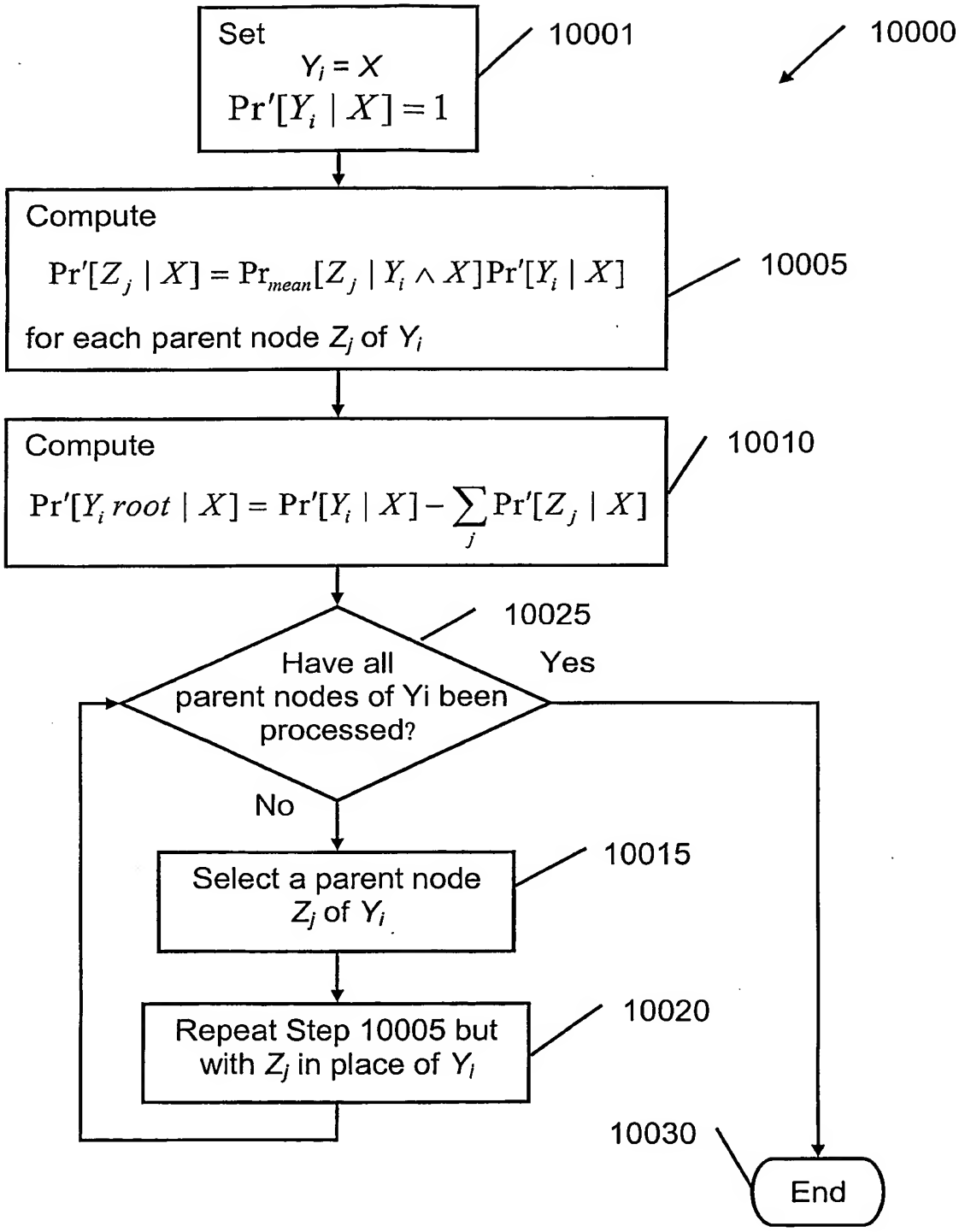
Set
$Y_i = X$
$\Pr'[Y_i \mid X] = 1$
10001
10000
Compute
$\Pr'[Z_j \mid X] = \Pr_{mean}[Z_j \mid Y_i \wedge X]\Pr'[Y_i \mid X]$
for each parent node Z_j of Y_i
10005
Compute
$\Pr'[Y_i\, root \mid X] = \Pr'[Y_i \mid X] - \sum_j \Pr'[Z_j \mid X]$
10010
10025
Have all parent nodes of Yi been processed?
Yes
No
Select a parent node Z_j of Y_i
10015
Repeat Step 10005 but with Z_j in place of Y_i
10020
10030
End

Fig. 10

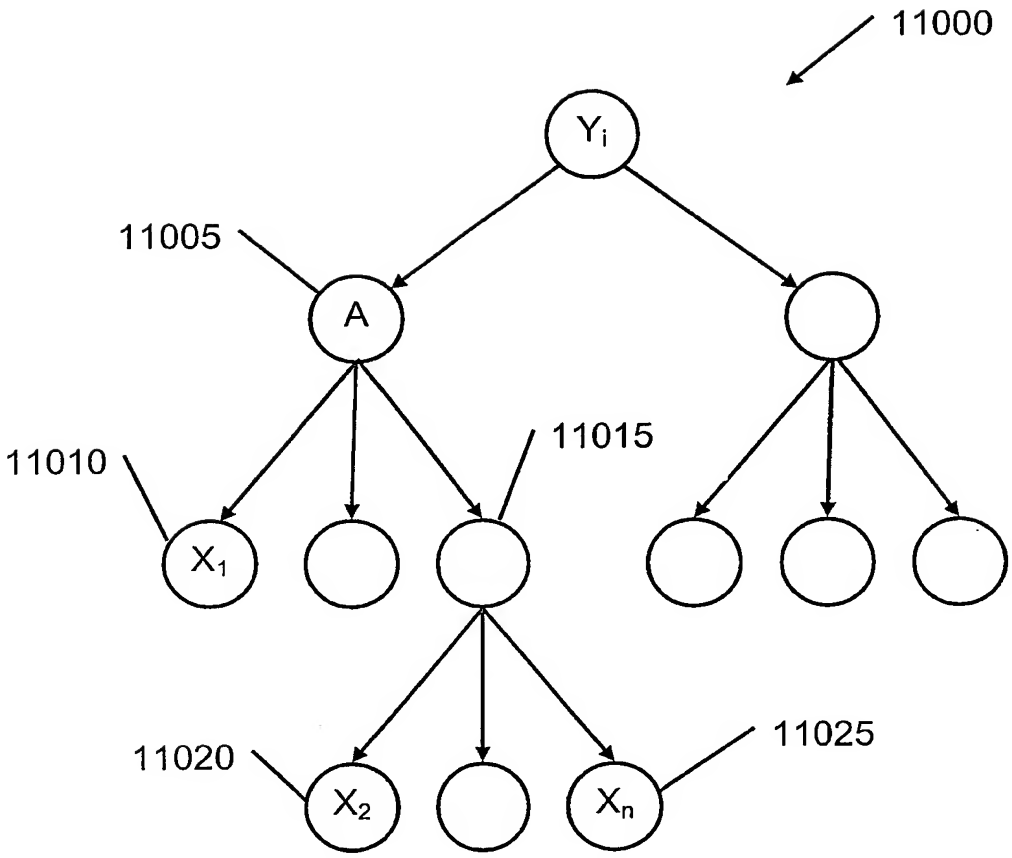

Fig. 11

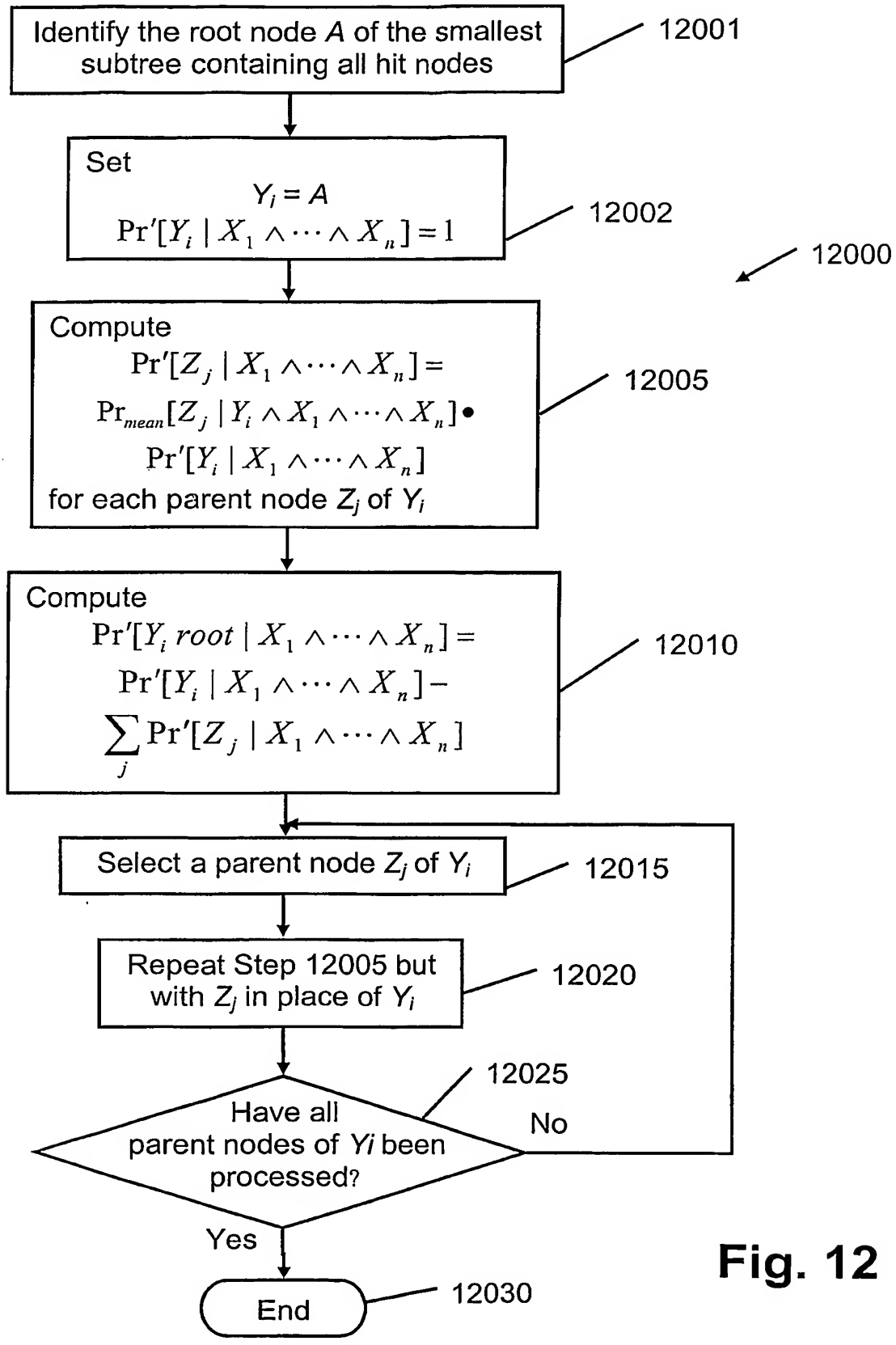
Identify the root node A of the smallest subtree containing all hit nodes
12001
Set
$Y_i = A$
$\Pr'[Y_i \mid X_1 \wedge \cdots \wedge X_n] = 1$
12002
12000
Compute
$\Pr'[Z_j \mid X_1 \wedge \cdots \wedge X_n] =$
$\Pr_{mean}[Z_j \mid Y_i \wedge X_1 \wedge \cdots \wedge X_n] \bullet$
$\Pr'[Y_i \mid X_1 \wedge \cdots \wedge X_n]$
for each parent node Z_j of Y_i
12005
Compute
$\Pr'[Y_i \; root \mid X_1 \wedge \cdots \wedge X_n] =$
$\Pr'[Y_i \mid X_1 \wedge \cdots \wedge X_n] -$
$\sum_j \Pr'[Z_j \mid X_1 \wedge \cdots \wedge X_n]$
12010
Select a parent node Z_j of Y_i
12015
Repeat Step 12005 but with Z_j in place of Y_i
12020
12025
Have all parent nodes of Yi been processed?
No
Yes
End
12030

Fig. 12

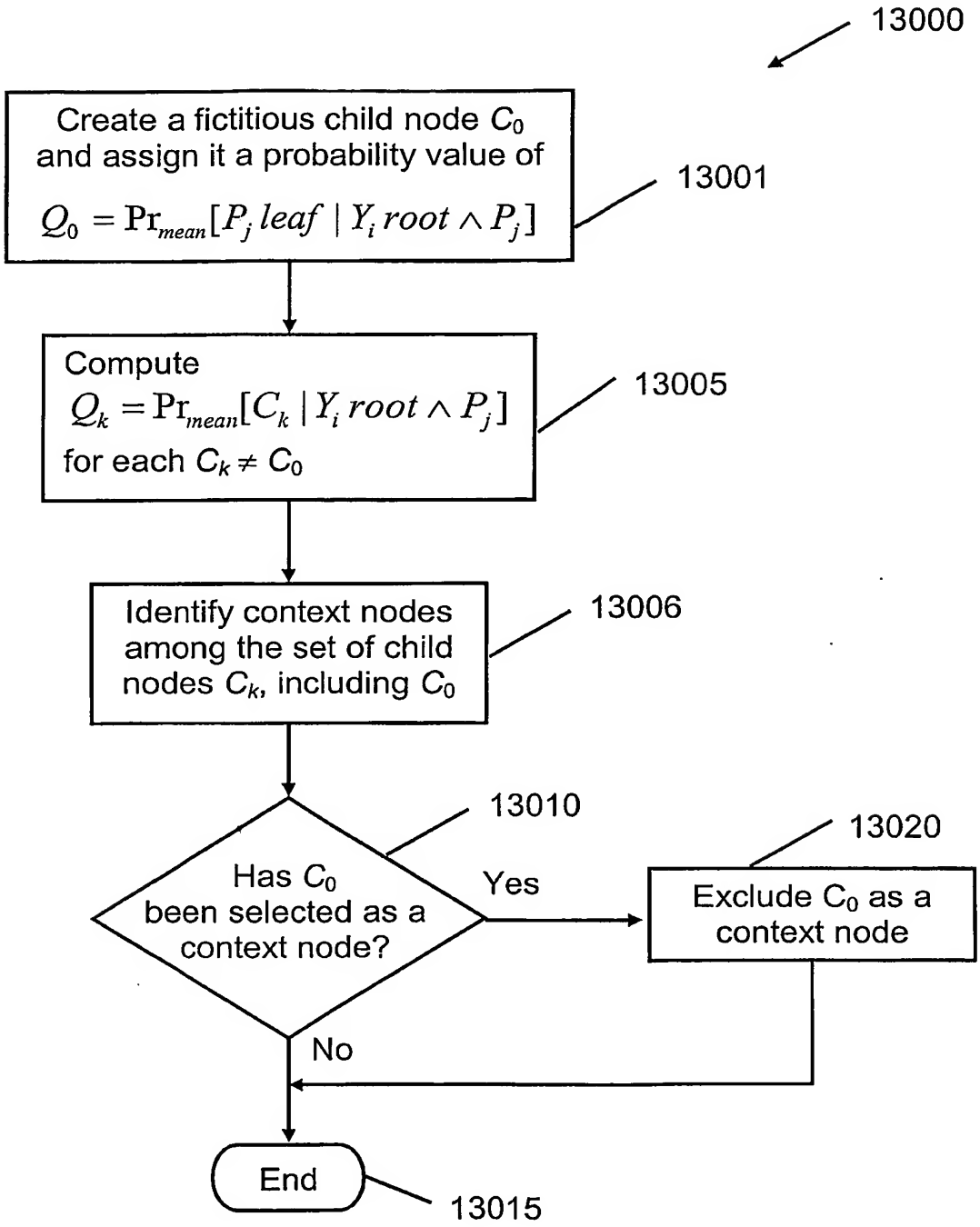
13000
Create a fictitious child node C_0 and assign it a probability value of $Q_0 = \Pr_{mean}[P_j\ leaf \mid Y_i\ root \wedge P_j]$
13001
Compute $Q_k = \Pr_{mean}[C_k \mid Y_i\ root \wedge P_j]$ for each $C_k \neq C_0$
13005
Identify context nodes among the set of child nodes C_k, including C_0
13006
Has C_0 been selected as a context node?
13010
Yes
Exclude C_0 as a context node
13020
No
End
13015

Fig. 13

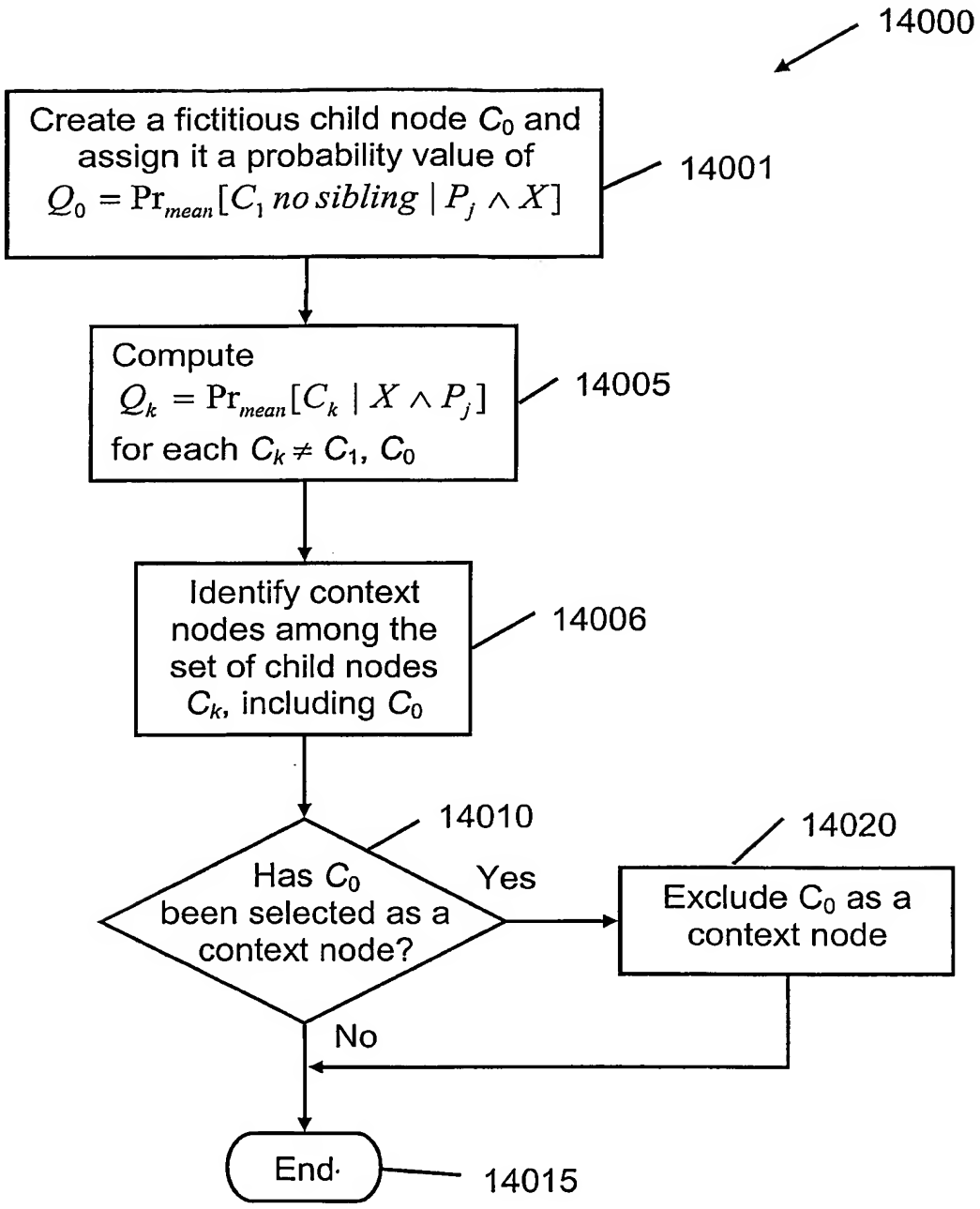
14000
Create a fictitious child node C_0 and assign it a probability value of $Q_0 = \Pr_{mean}[C_1 \, no \, sibling \mid P_j \wedge X]$
14001
Compute $Q_k = \Pr_{mean}[C_k \mid X \wedge P_j]$ for each $C_k \neq C_1, C_0$
14005
Identify context nodes among the set of child nodes C_k, including C_0
14006
14010
Has C_0 been selected as a context node?
Yes
14020
Exclude C_0 as a context node
No
End
14015

Fig. 14

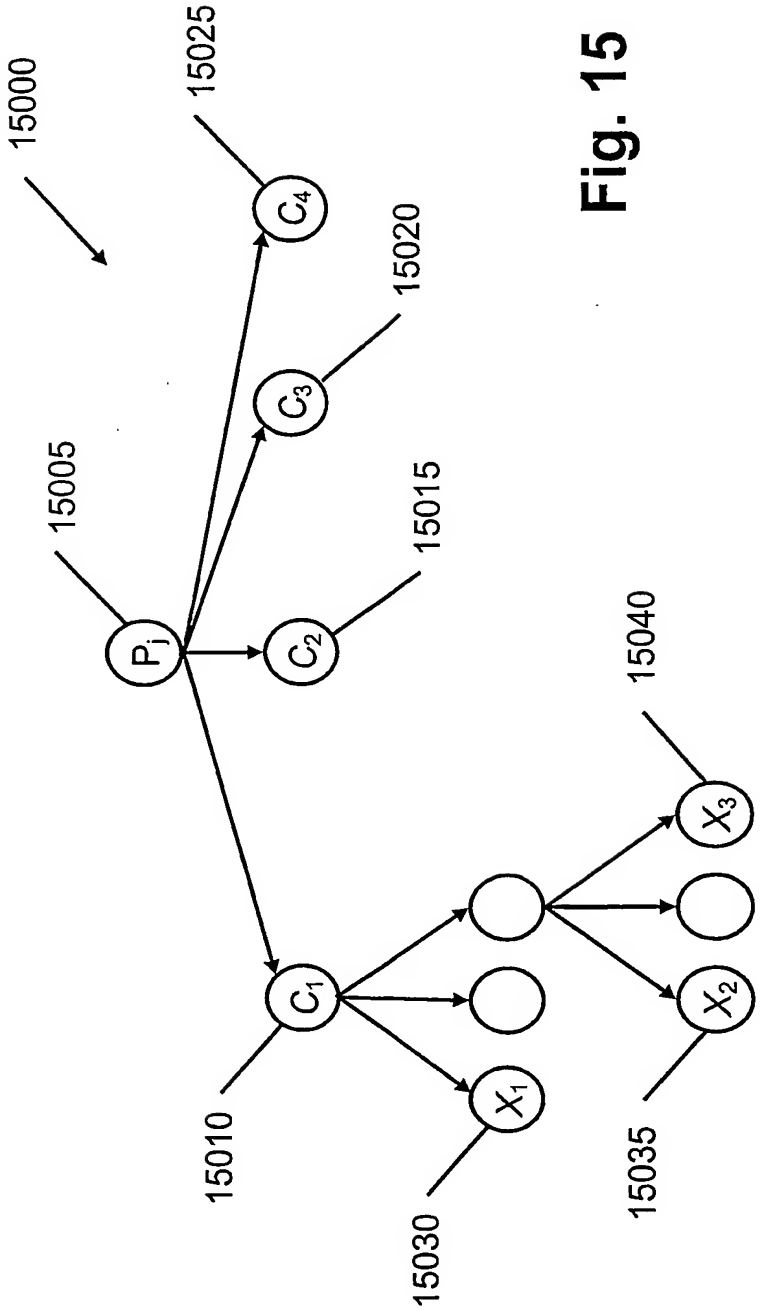
15000
15025
15020
15005
15015
15040
15010
15030
15035
P_j
C_1
C_2
C_3
C_4
X_1
X_2
X_3

Fig. 15

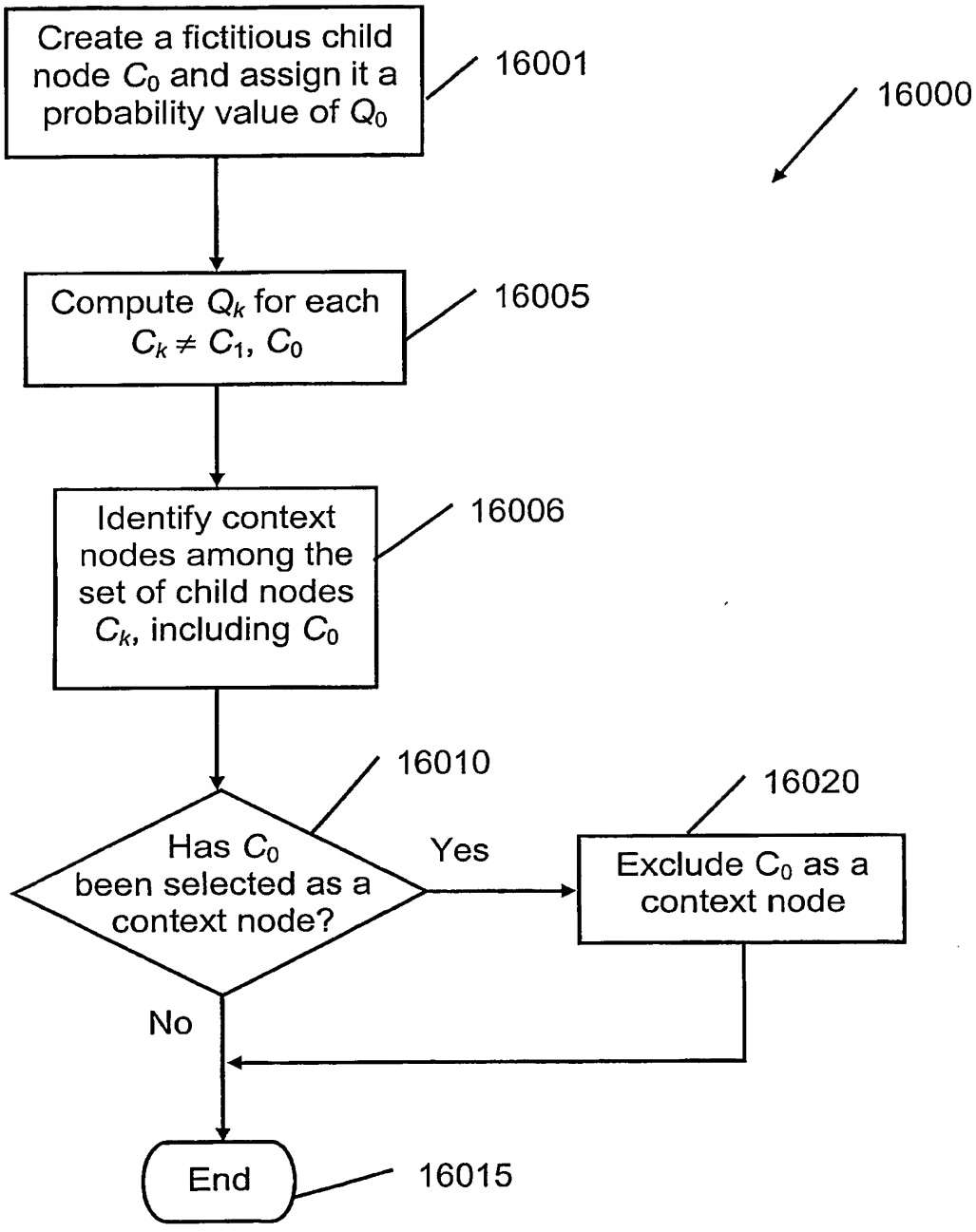
Create a fictitious child node C_0 and assign it a probability value of Q_0
16001
16000
Compute Q_k for each $C_k \neq C_1, C_0$
16005
Identify context nodes among the set of child nodes C_k, including C_0
16006
16010
Has C_0 been selected as a context node?
Yes
16020
Exclude C_0 as a context node
No
End
16015

Fig. 16

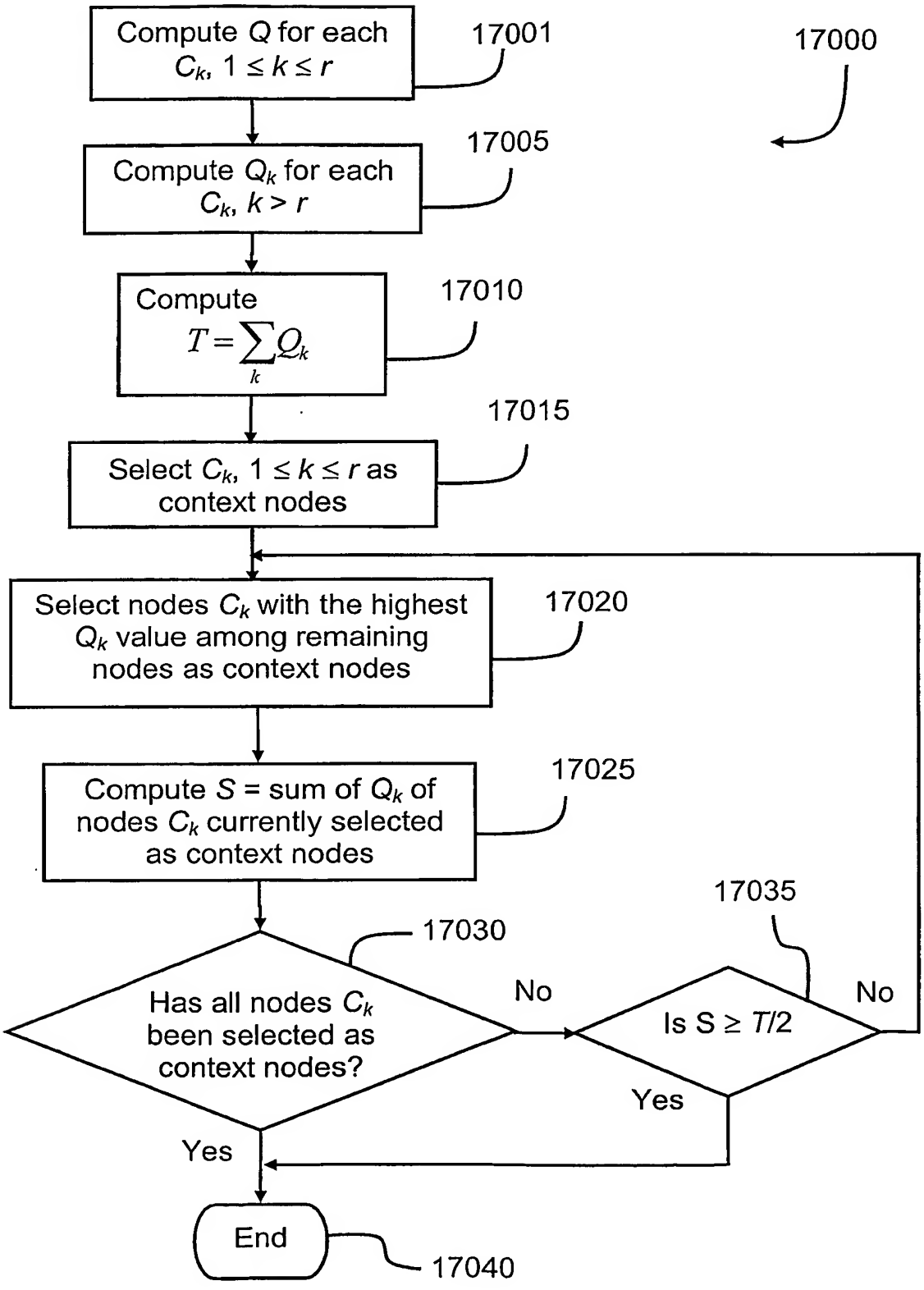
Compute Q for each C_k, $1 \le k \le r$
17001
17000
Compute Q_k for each C_k, $k > r$
17005
Compute $T = \sum_k Q_k$
17010
Select C_k, $1 \le k \le r$ as context nodes
17015
Select nodes C_k with the highest Q_k value among remaining nodes as context nodes
17020
Compute S = sum of Q_k of nodes C_k currently selected as context nodes
17025
17035
17030
Has all nodes C_k been selected as context nodes?
No
Is $S \ge T/2$
No
Yes
Yes
End
17040

Fig. 17

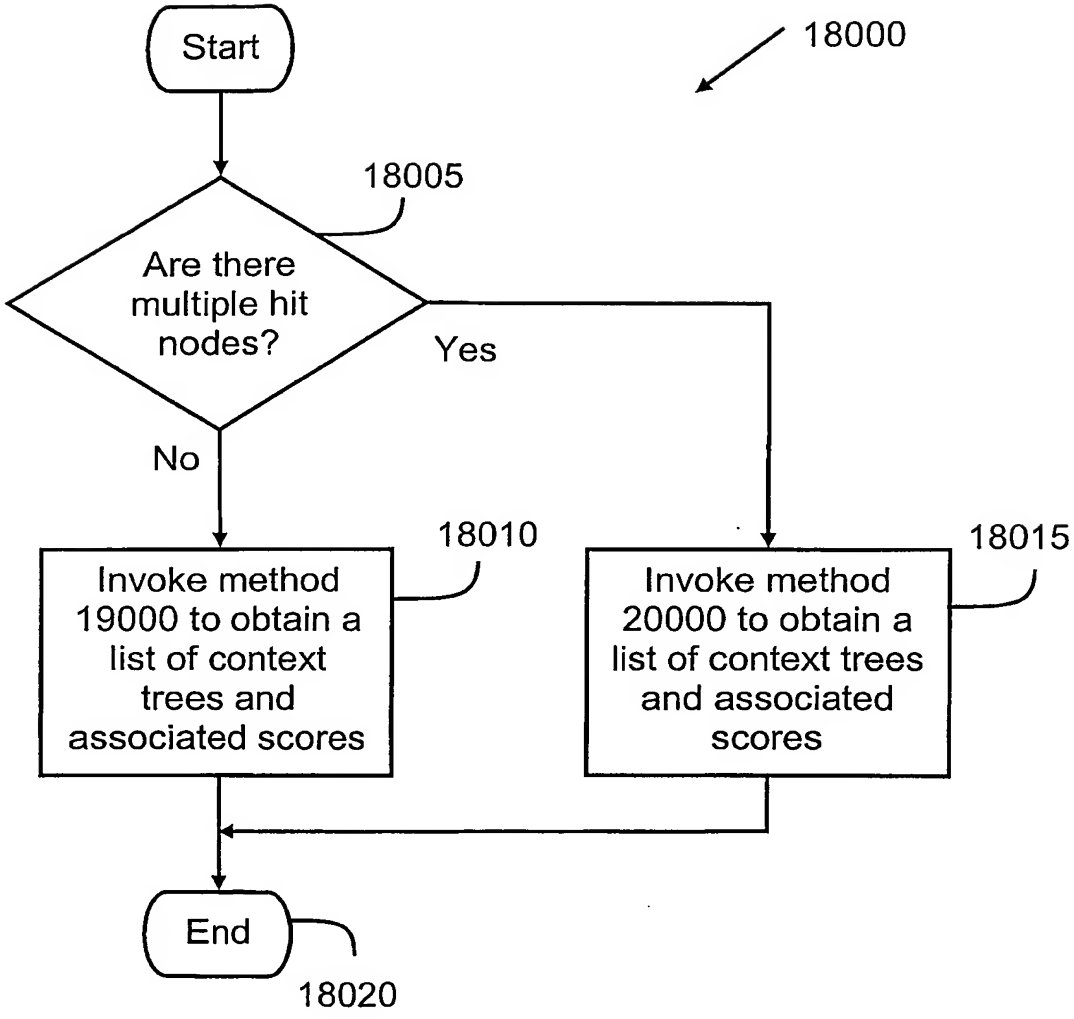
Start
18000
18005
Are there multiple hit nodes?
Yes
No
18010
Invoke method 19000 to obtain a list of context trees and associated scores
18015
Invoke method 20000 to obtain a list of context trees and associated scores
End
18020

Fig. 18

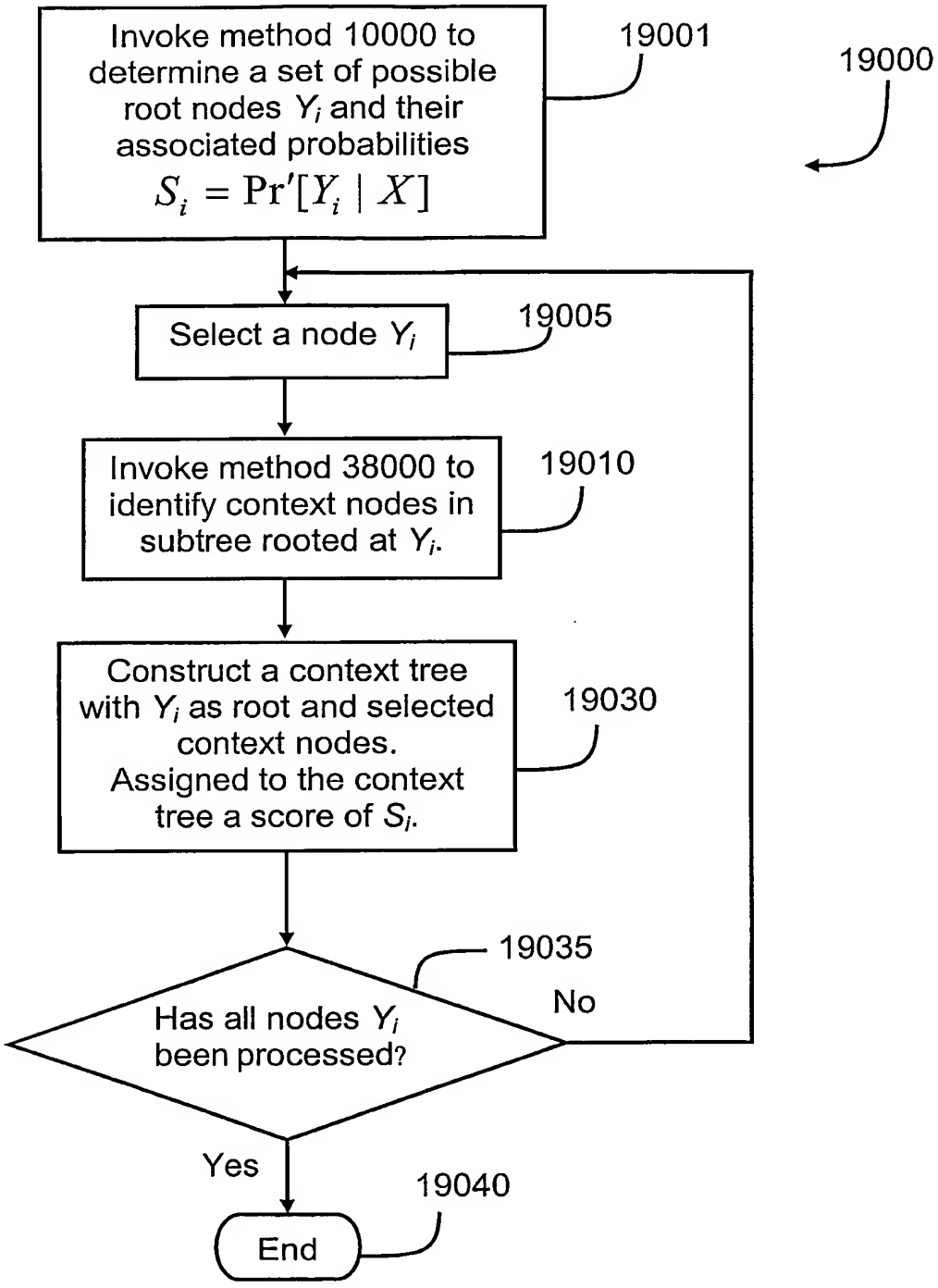
Invoke method 10000 to determine a set of possible root nodes Y_i and their associated probabilities $S_i = \Pr'[Y_i \mid X]$
19001
19000
Select a node Y_i
19005
Invoke method 38000 to identify context nodes in subtree rooted at Y_i.
19010
Construct a context tree with Y_i as root and selected context nodes. Assigned to the context tree a score of S_i.
19030
Has all nodes Y_i been processed?
19035
No
Yes
End
19040

Fig. 19

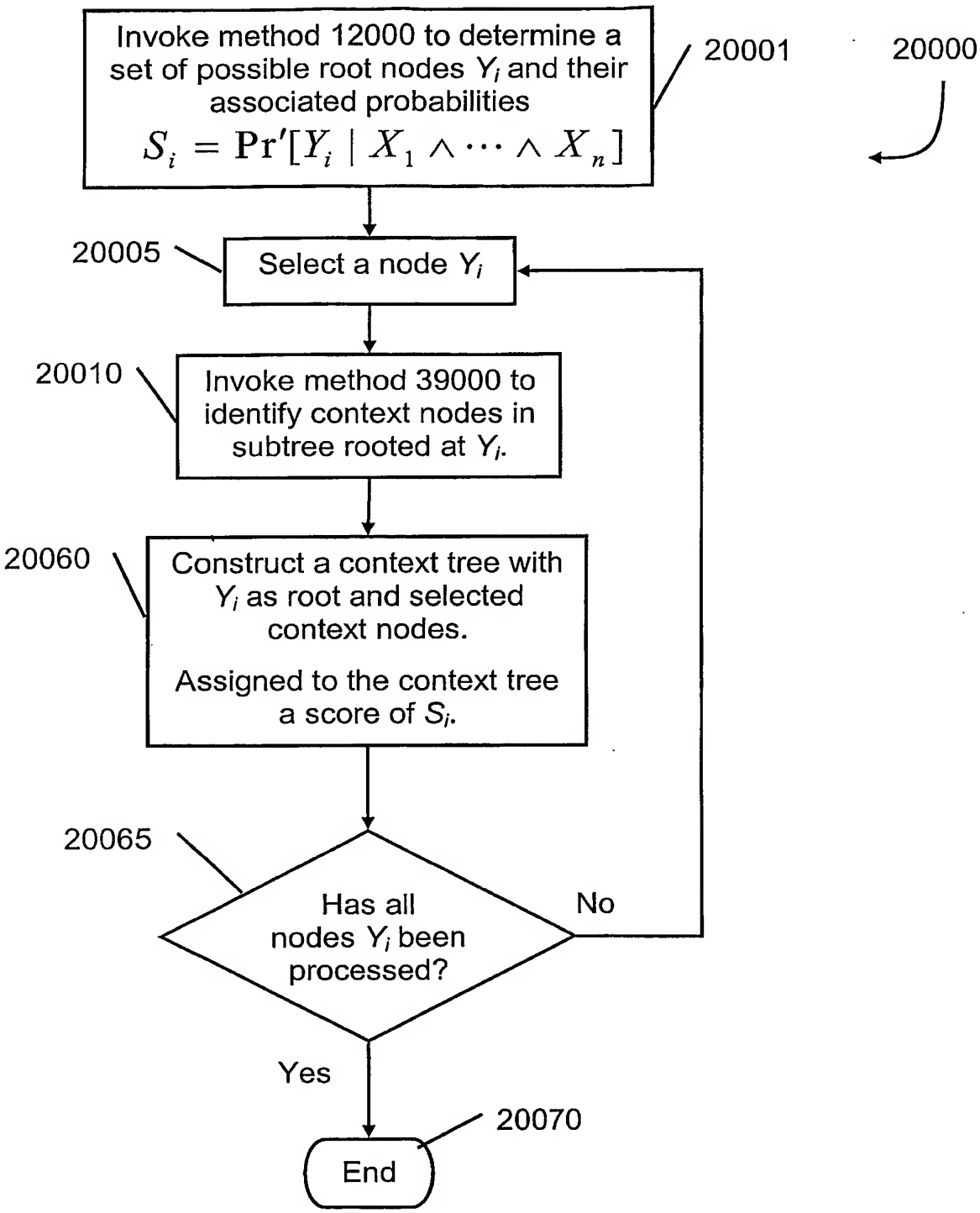
Invoke method 12000 to determine a set of possible root nodes Y_i and their associated probabilities
$S_i = \Pr'[Y_i \mid X_1 \wedge \cdots \wedge X_n]$
20001
20000
20005
Select a node Y_i
20010
Invoke method 39000 to identify context nodes in subtree rooted at Y_i.
20060
Construct a context tree with Y_i as root and selected context nodes.
Assigned to the context tree a score of S_i.
20065
Has all nodes Y_i been processed?
No
Yes
20070
End

Fig. 20

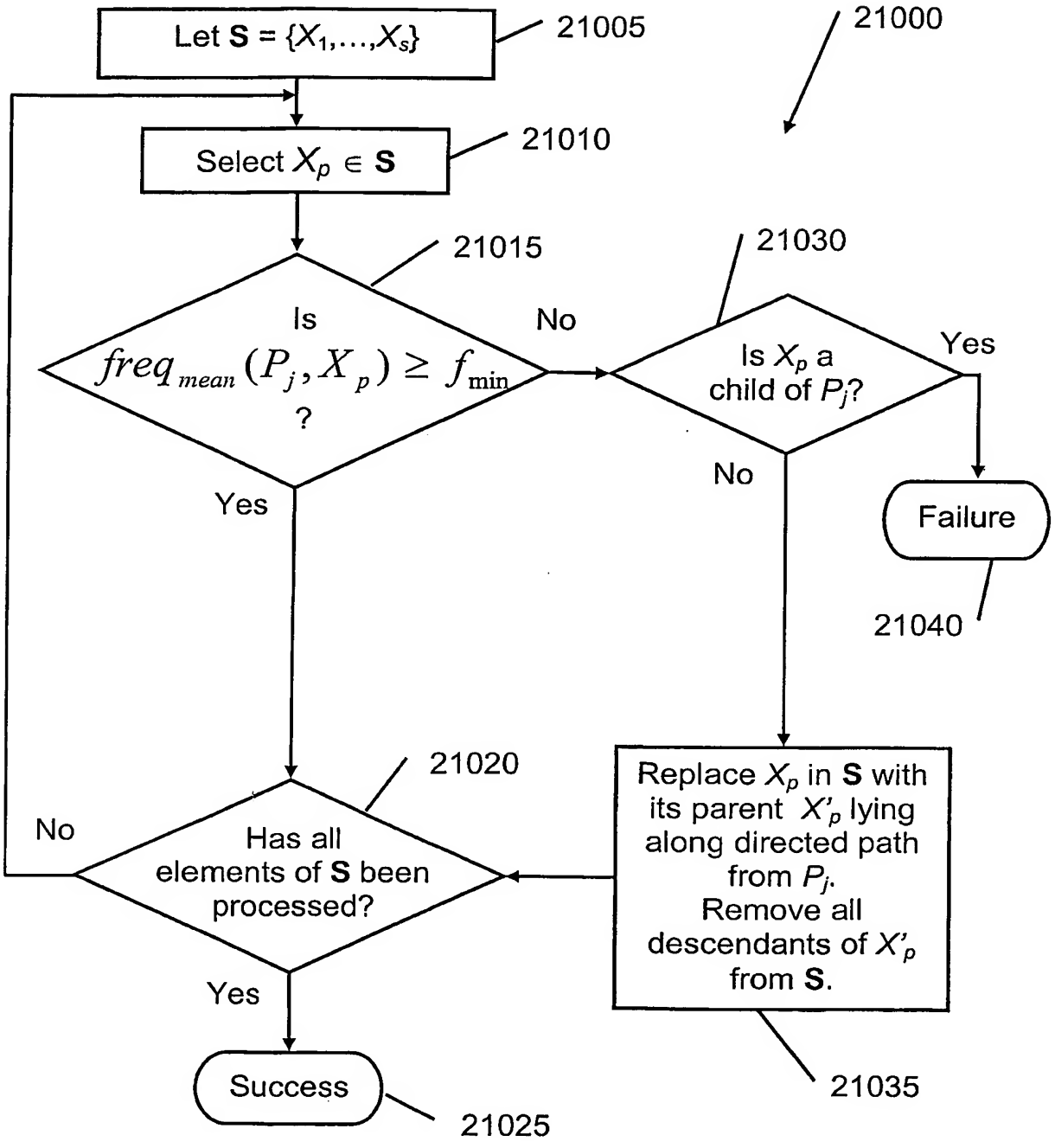
Let S = {X1,...,Xs}
21005
21000
Select Xp ∈ S
21010
21015
21030
Is freq mean (Pj, Xp) ≥ f min ?
No
Is Xp a child of Pj?
Yes
No
Yes
Failure
21040
21020
No
Has all elements of S been processed?
Replace Xp in S with its parent X'p lying along directed path from Pj. Remove all descendants of X'p from S.
Yes
Success
21025
21035

Fig. 21

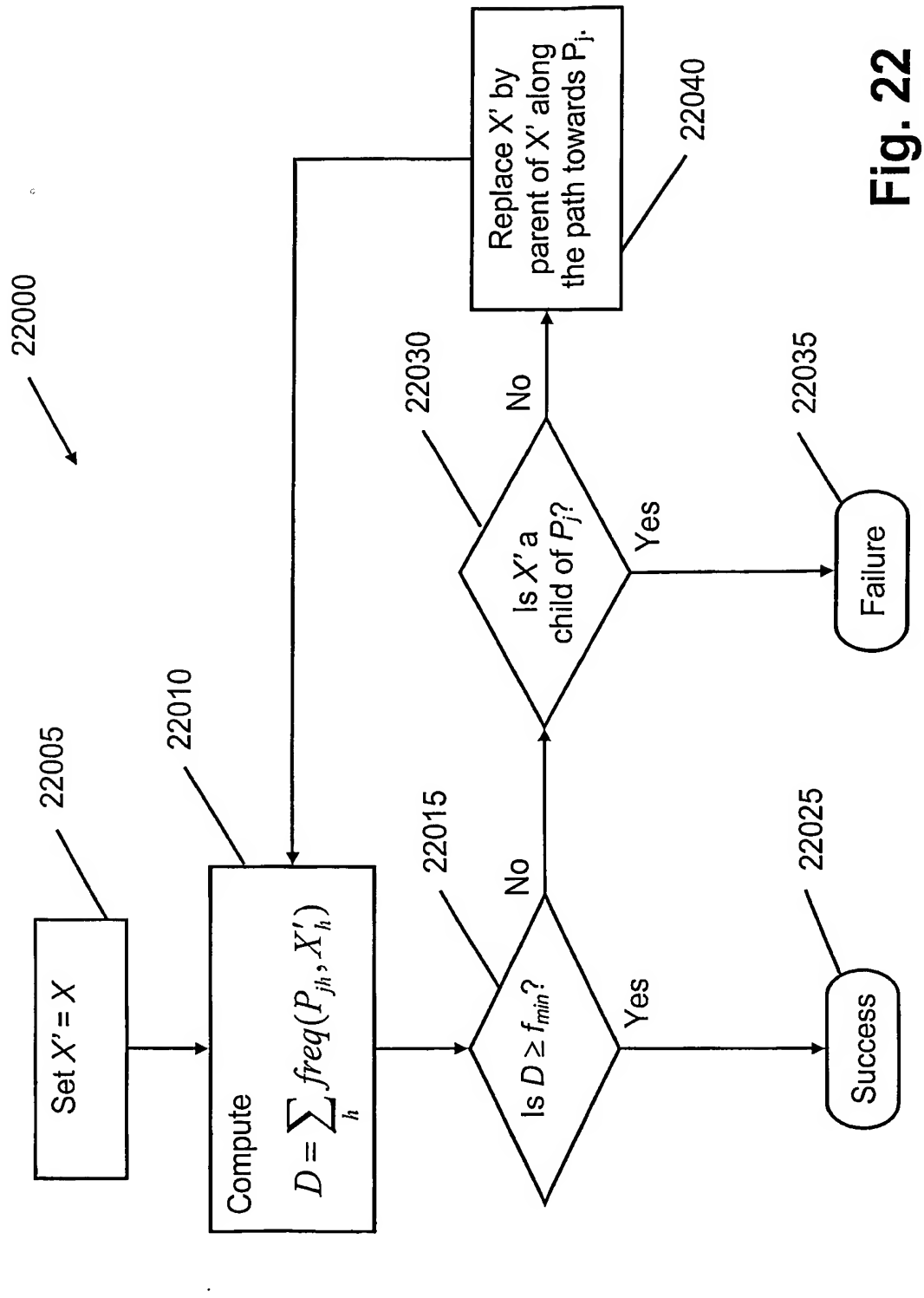
22000
22005
Set X' = X
22010
Compute
$D = \sum_h freq(P_{jh}, X'_h)$
22015
Is $D \geq f_{min}$?
No
Yes
22025
Success
22030
Is X' a child of P_j?
No
Yes
22035
Failure
22040
Replace X' by parent of X' along the path towards P_j.

Fig. 22

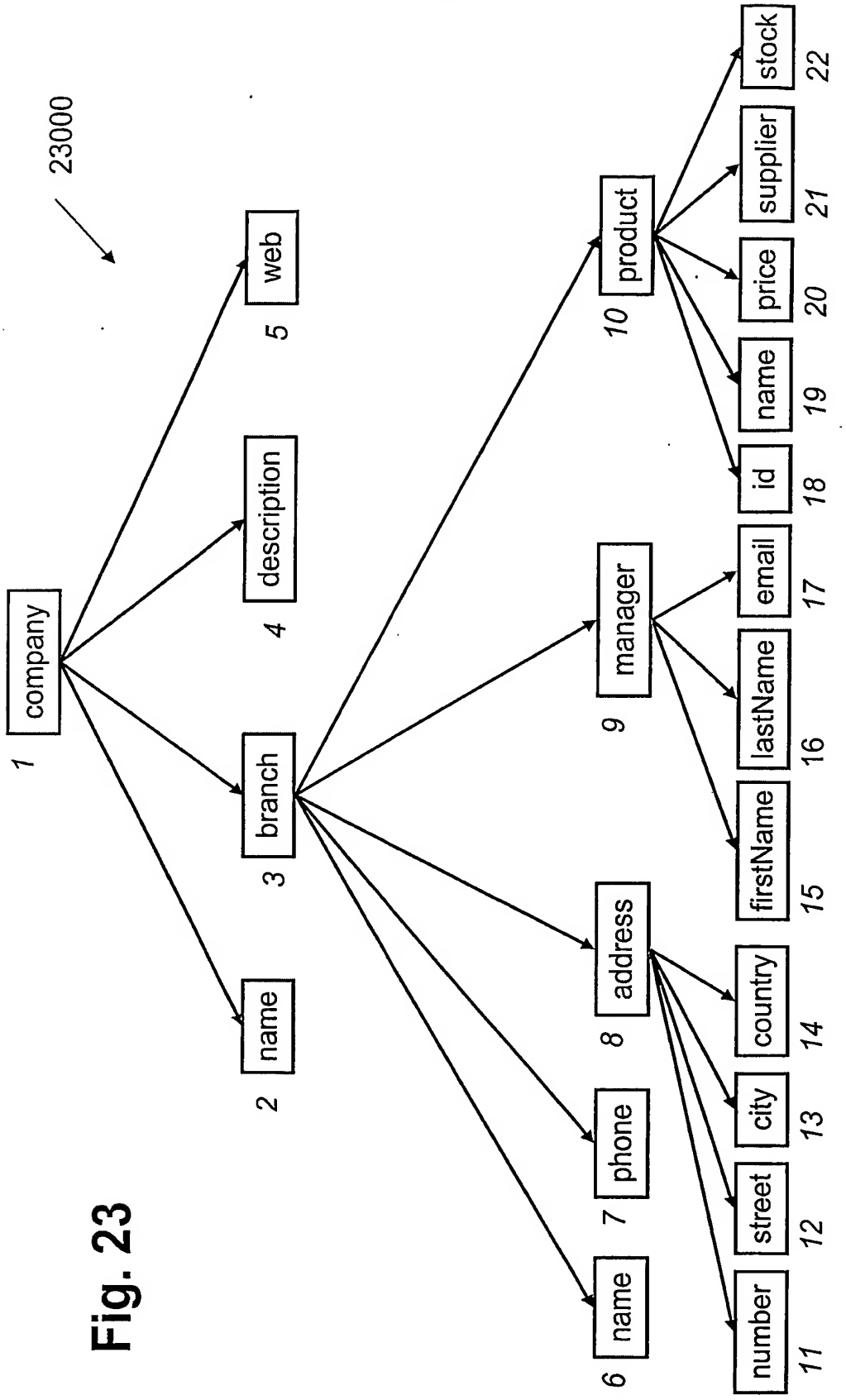
23000
1 company
2 name
3 branch
4 description
5 web
6 name
7 phone
8 address
9 manager
10 product
11 number
12 street
13 city
14 country
15 firstName
16 lastName
17 email
18 id
19 name
20 price
21 supplier
22 stock

Fig. 23

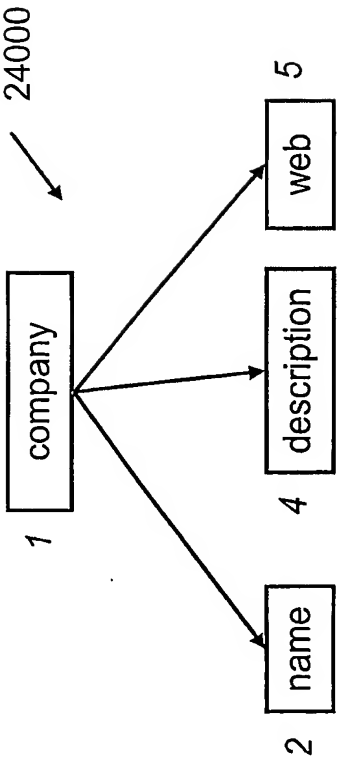
24000
company
1
name
2
description
4
web
5

Fig. 24

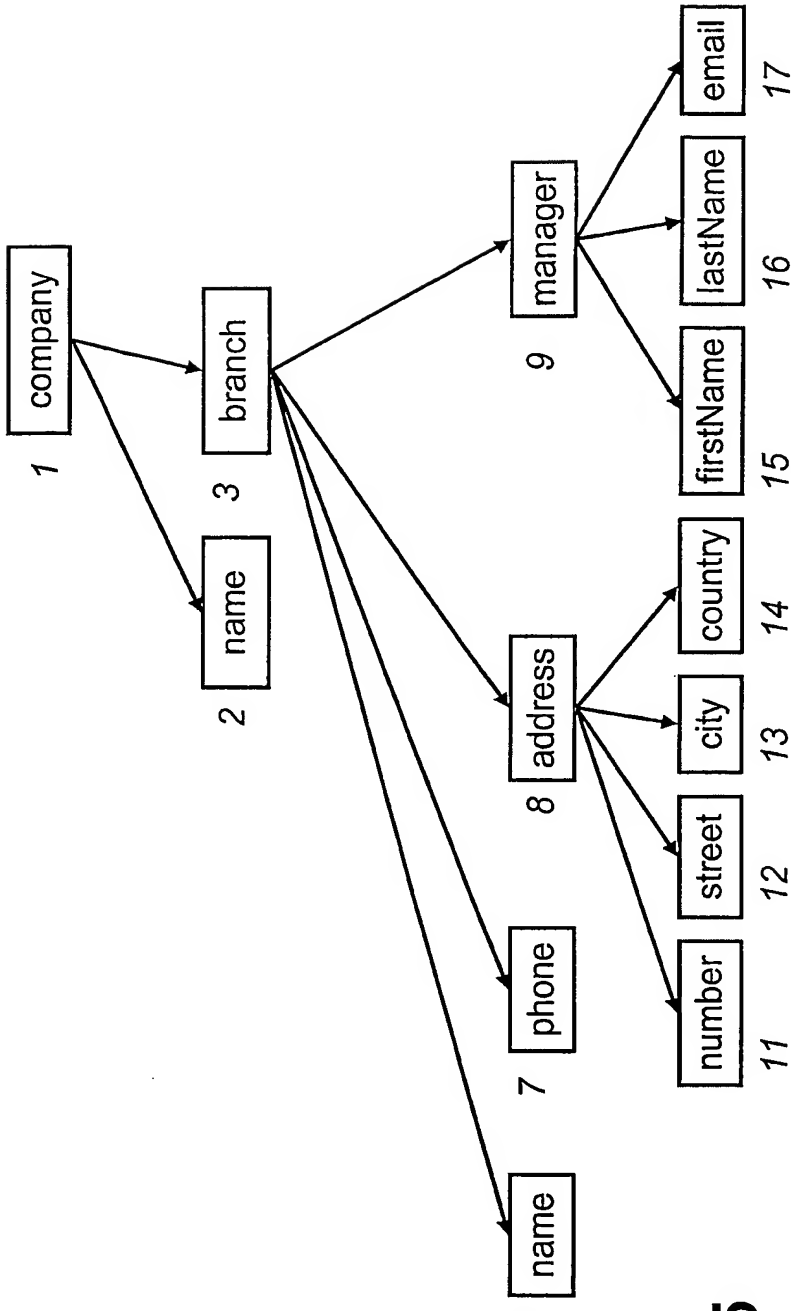
25000
company
1
name
2
branch
3
name
6
phone
7
address
8
manager
9
number
11
street
12
city
13
country
14
firstName
15
lastName
16
email
17

Fig. 25

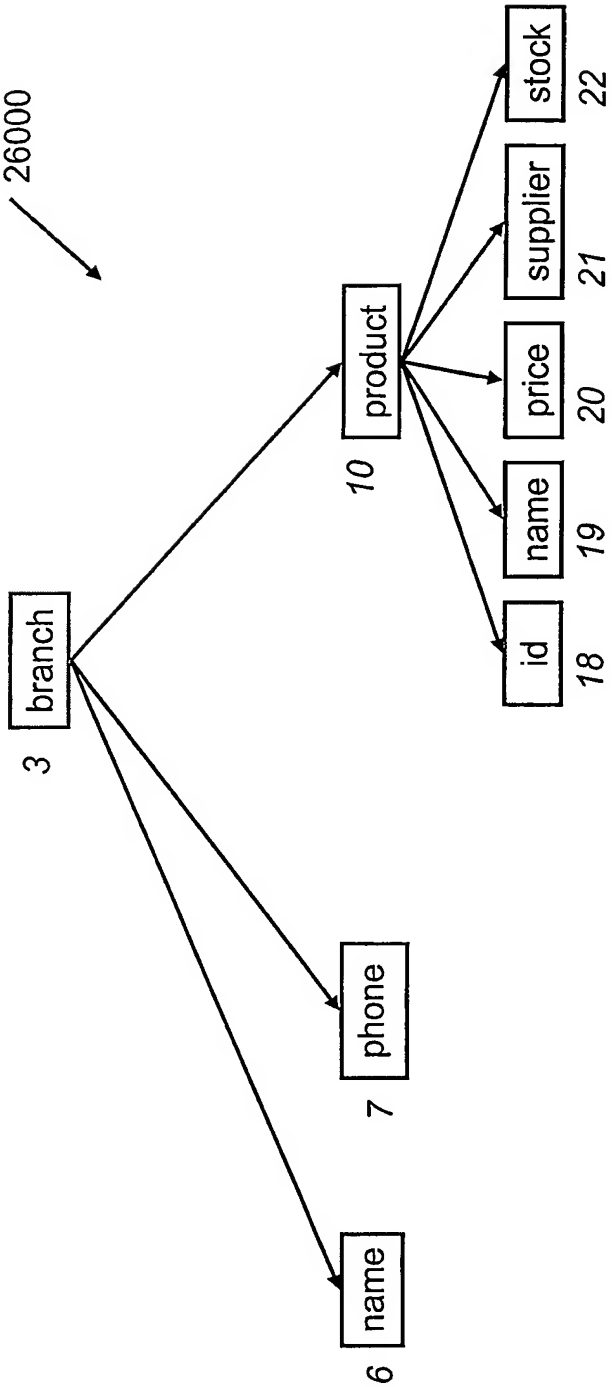
26000
branch
3
product
10
id
18
name
19
price
20
supplier
21
stock
22
phone
7
name
6

Fig. 26

X	*freq(X)*
1	2
2	2
3	2
4	1
5	1
6	2
7	2
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1

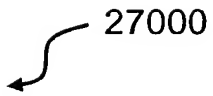

Fig. 27

freq(Y_i, X)

28000

	Y_i				
X	1	3	8	9	10
2	2				
3	1				
4	1				
5	1				
6	1	2			
7	1	2			
8	1	1			
9	1	1			
10	0	1			
11	1	1	1		
12	1	1	1		
13	1	1	1		
14	1	1	1		
15	1	1		1	
16	1	1		1	
17	1	1		1	
18	0	1			1
19	0	1			1
20	0	1			1
21	0	1			1
22	0	1			1

28005

Fig. 28

freq(Y_i, P_j *leaf*)

	Y_i	
P_j	1	3
3	0	
8	0	0
9	0	0
10	0	0

29000

Fig. 29

freq(P_j has 1 child, X)

	P_j				
X	1	3	8	9	10
2	0				
3	0				
4	0				
5	0				
6	0	0			
7	0	0			
8	0	0			
9	0	0			
10	0	0			
11	0	0	0		
12	0	0	0		
13	0	0	0		
14	0	0	0		
15	0	0		0	
16	0	0		0	
17	0	0		0	
18	0	0			0
19	0	0			0
20	0	0			0
21	0	0			0
22	0	0			0

30000

Fig. 30

$freq(C_k, P_j, X)$: P_j = **node 1**

31000

	C_k			
X	2	3	4	5
2		1	1	1
3	1		0	0
4	1	0		1
5	1	0	1	
6	1		0	0
7	1		0	0
8	1		0	0
9	1		0	0
10	0		0	0
11	1		0	0
12	1		0	0
13	1		0	0
14	1		0	0
15	1		0	0
16	1		0	0
17	1		0	0
18	0		0	0
19	0		0	0
20	0		0	0
21	0		0	0
22	0		0	0

Fig. 31

***freq*(C_k, P_j, X): P_j = node 3**

	C_k				
X	6	7	8	9	10
6		2	1	1	1
7	2		1	1	1
8	1	1		1	0
9	1	1	1		0
10	1	1	0	0	
11	1	1		1	0
12	1	1		1	0
13	1	1		1	0
14	1	1		1	0
15	1	1	1		0
16	1	1	1		0
17	1	1	1		0
18	1	1	0	0	
19	1	1	0	0	
20	1	1	0	0	
21	1	1	0	0	
22	1	1	0	0	

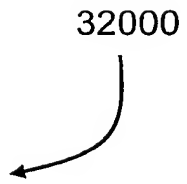

Fig. 32

***freq*(C_k, P_j, X): P_j = node 8**

33000

X	C_k 11	12	13	14
11		1	1	1
12	1		1	1
13	1	1		1
14	1	1	1	

Fig. 33

***freq*(C_k, P_j, X): P_j = node 9**

34000

X	C_k 15	16	17
15		1	1
16	1		1
17	1	1	

Fig. 34

***freq*(C_k, P_j, X): P_j = node 10**

35000

X	C_k 18	19	20	21	22
18		1	1	1	1
19	1		1	1	1
20	1	1		1	1
21	1	1	1		1
22	1	1	1	1	

Fig. 35

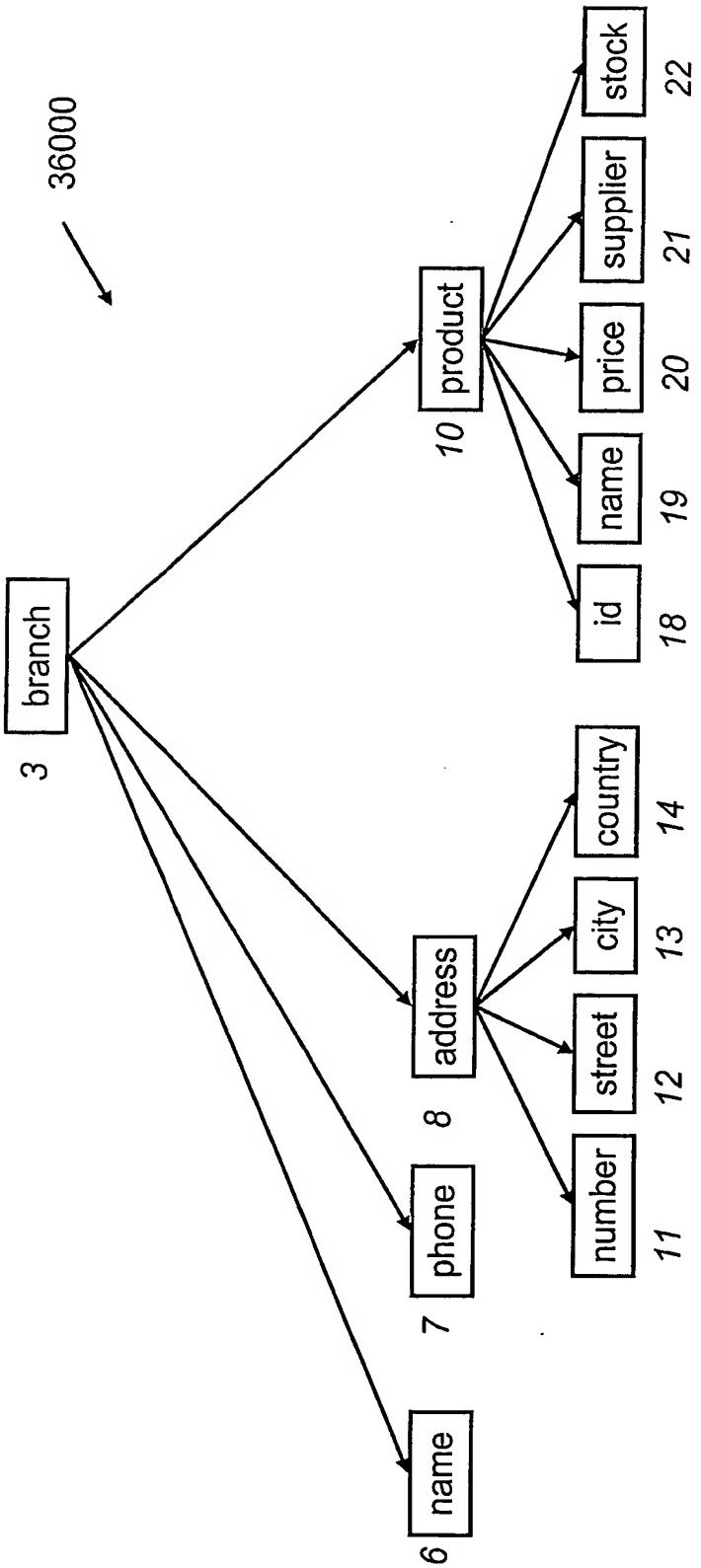
36000
branch
3
product
10
id
18
name
19
price
20
supplier
21
stock
22
address
8
country
14
city
13
street
12
number
11
phone
7
name
6

Fig. 36

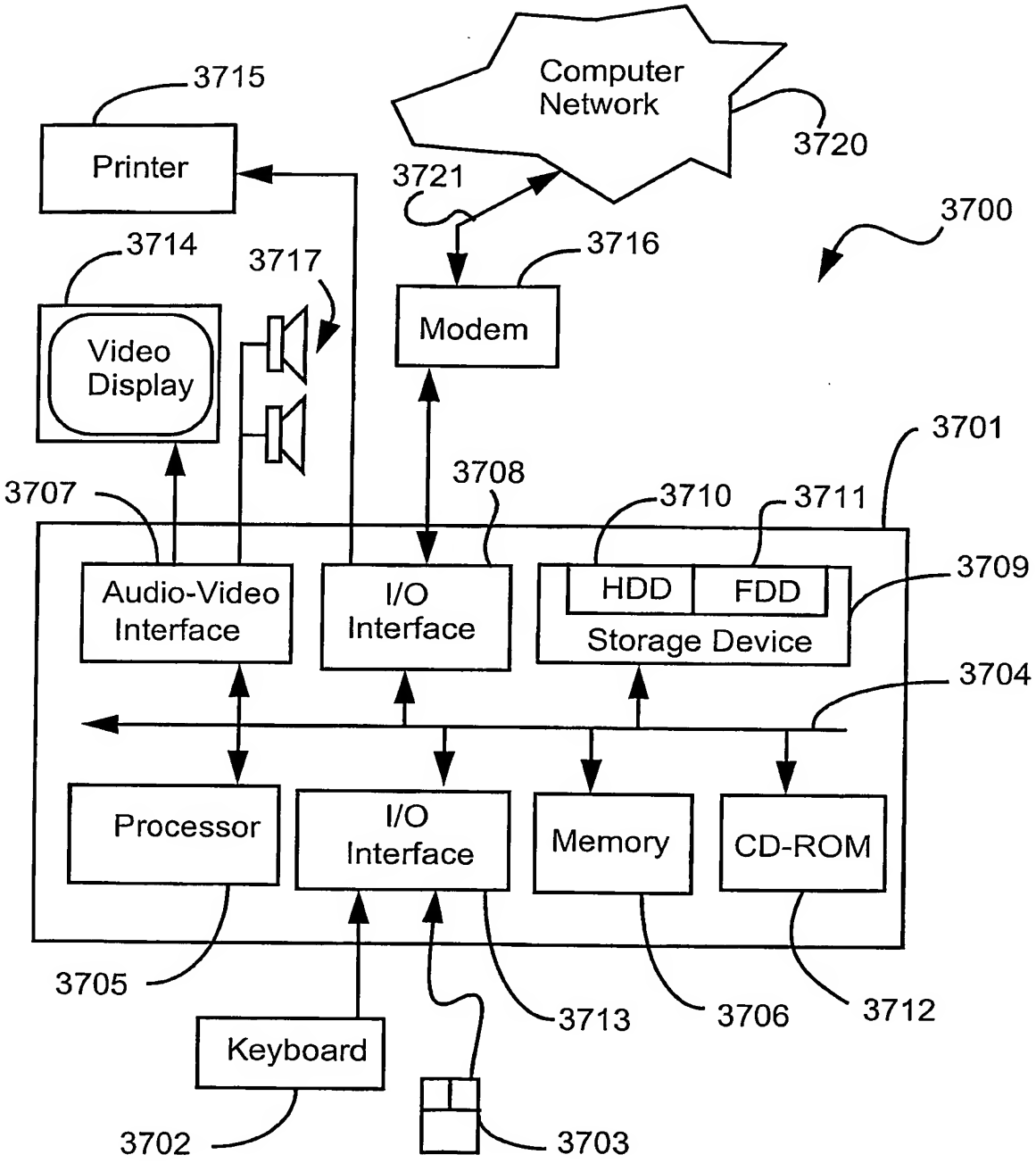
Computer Network
3720
3715
Printer
3721
3700
3714
3717
3716
Video Display
Modem
3701
3708
3707
3710
3711
3709
Audio-Video Interface
I/O Interface
HDD
FDD
Storage Device
3704
Processor
I/O Interface
Memory
CD-ROM
3705
3713
3706
3712
Keyboard
3702
3703

Fig. 37

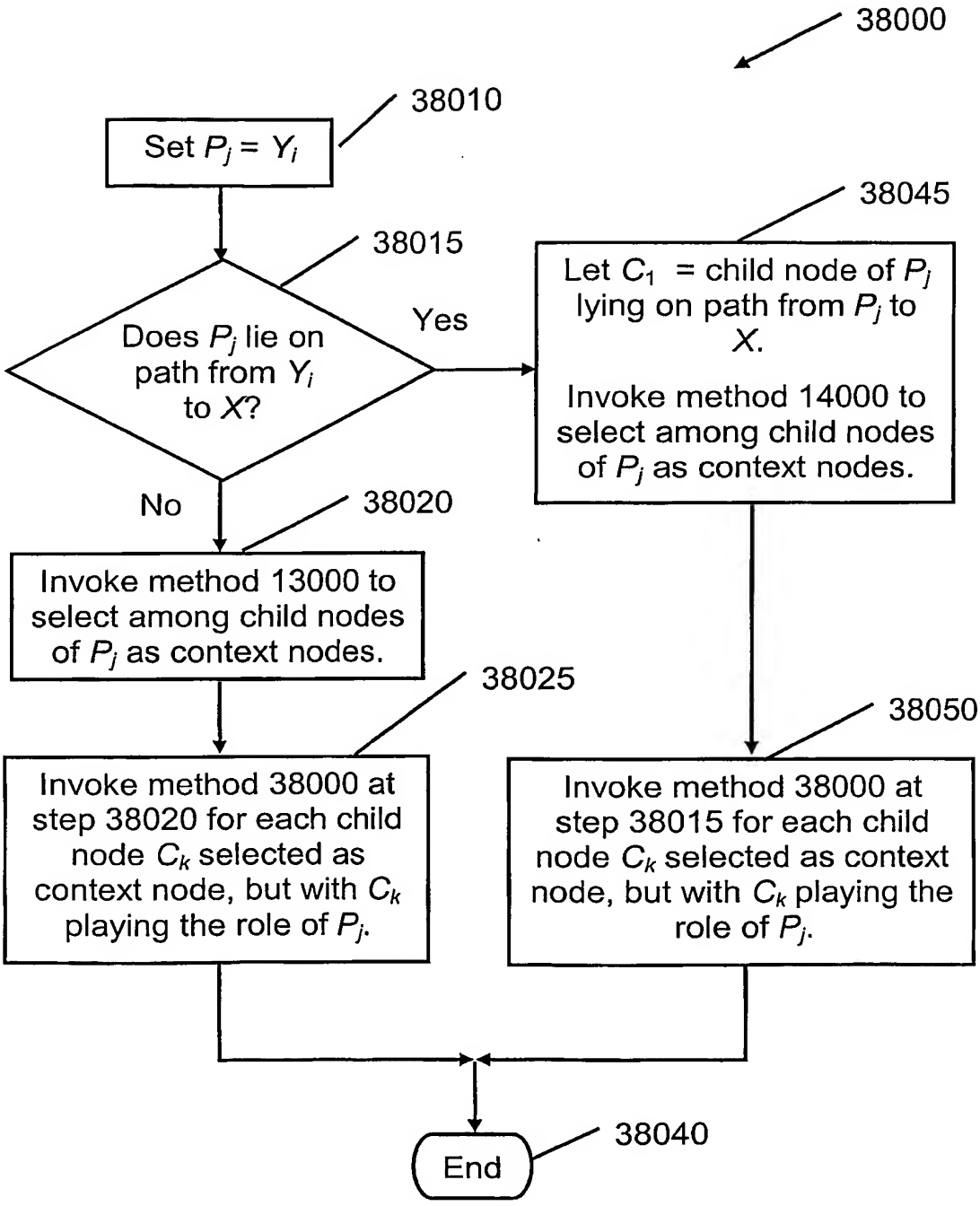
38000
38010
Set $P_j = Y_i$
38015
Does P_j lie on path from Y_i to X?
Yes
No
38045
Let C_1 = child node of P_j lying on path from P_j to X.
Invoke method 14000 to select among child nodes of P_j as context nodes.
38020
Invoke method 13000 to select among child nodes of P_j as context nodes.
38025
Invoke method 38000 at step 38020 for each child node C_k selected as context node, but with C_k playing the role of P_j.
38050
Invoke method 38000 at step 38015 for each child node C_k selected as context node, but with C_k playing the role of P_j.
38040
End

Fig. 38

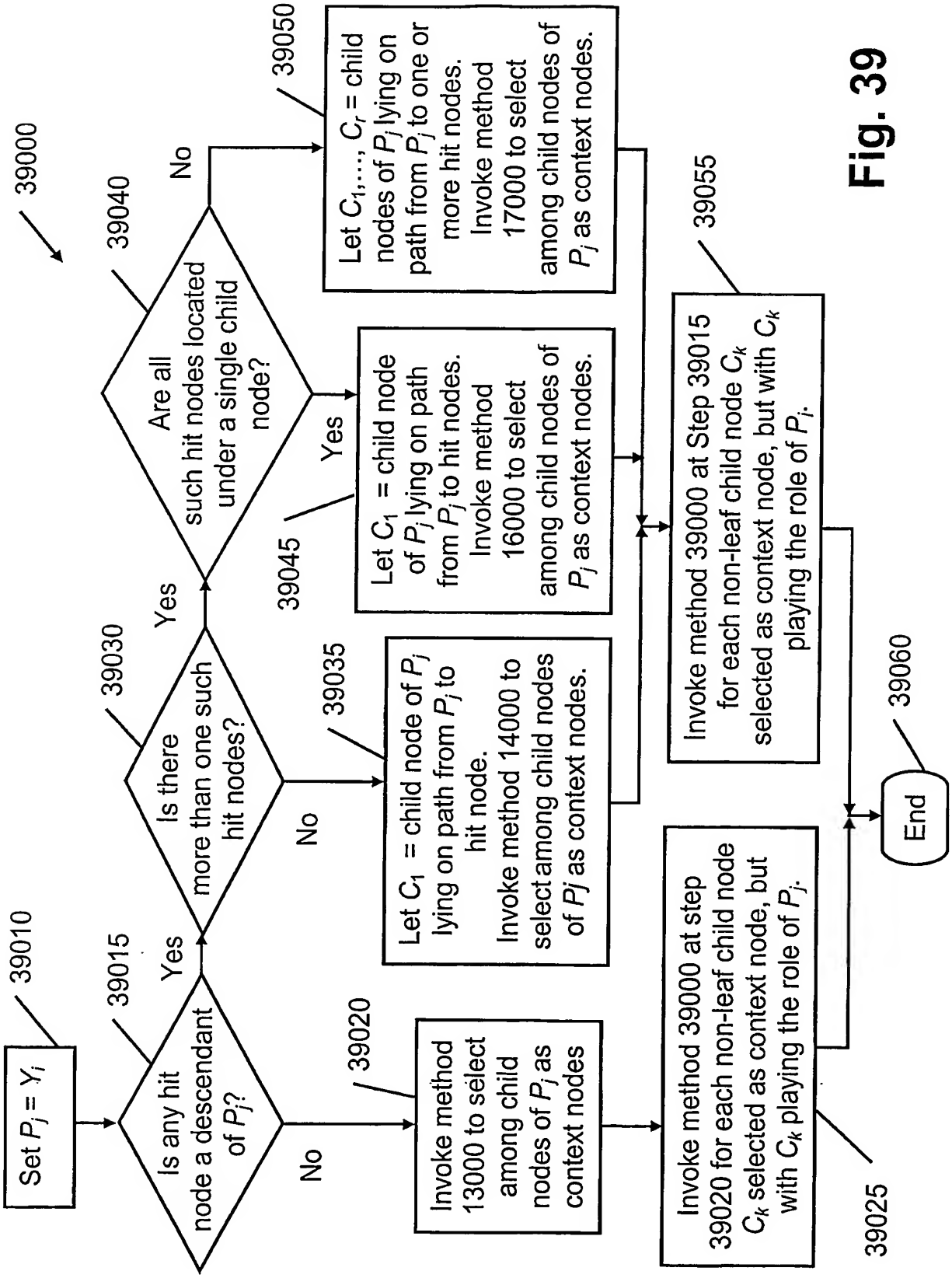
39000
Set $P_j = Y_i$
39010
Is any hit node a descendant of P_j?
39015
Yes
No
Invoke method 13000 to select among child nodes of P_j as context nodes
39020
Is there more than one such hit nodes?
39030
Yes
No
Let C_1 = child node of P_j lying on path from P_j to hit node. Invoke method 14000 to select among child nodes of Pj as context nodes.
39035
Are all such hit nodes located under a single child node?
39040
Yes
No
Let C_1 = child node of P_j lying on path from P_j to hit nodes. Invoke method 16000 to select among child nodes of P_j as context nodes.
39045
Let $C_1,\ldots, C_r$ = child nodes of P_j lying on path from P_j to one or more hit nodes. Invoke method 17000 to select among child nodes of P_j as context nodes.
39050
Invoke method 39000 at step 39020 for each non-leaf child node C_k selected as context node, but with C_k playing the role of P_j.
39025
Invoke method 39000 at Step 39015 for each non-leaf child node C_k selected as context node, but with C_k playing the role of P_j.
39055
End
39060

Fig. 39